
VITAL EMO: EL DETECTOR DE ABURRIMIENTO DESDE UNA PERSPECTIVA DE MACHINE LEARNING

VITAL EMO: The boredom detector from a machine learning perspective.

DIEGO ISAR MUÑOZ
ÁLVARO DAVID ORTIZ MARCHUT
DANIEL PADILLA RODRÍGUEZ
SOFÍA PRIETO IBÁÑEZ



**UNIVERSIDAD COMPLUTENSE
MADRID**

Ingeniería del Software
FACULTAD DE INFORMÁTICA

Dirigido por:
Borja Manero
Meriem El Yamri

MADRID, 2020–2021

VITAL EMO

EL DETECTOR DE ABURRIMIENTO DESDE UNA PERSPECTIVA DE MACHINE LEARNING

Memoria que se presenta para el Trabajo de Fin de Grado

DIEGO ISAR MUÑOZ
ÁLVARO DAVID ORTIZ MARCHUT
DANIEL PADILLA RODRÍGUEZ
SOFÍA PRIETO IBÁÑEZ

Dirigido por

Borja Manero Iglesias
Meriem El Yamri

Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid

Madrid, 2021

Agradecimientos

Este proyecto no habría sido posible sin el apoyo de un gran número de personas:

En primer lugar, queremos agradecer a Josefa Ros Velasco por su guía en el ámbito del aburrimiento, proporcionándonos material de estudio, aclarando amablemente nuestras dudas y contribuyendo activamente en nuestra investigación.

Con especial cariño, queremos mencionar a nuestros familiares, amigos y compañeros. Por regalarnos su valioso tiempo al participar activamente en la creación del dataset, proporcionarnos apoyo en este año tan complicado y además ayudarnos siempre a seguir adelante con nuestras ideas.

Por último, queremos agradecer a nuestros tutores del TFG, Borja Manero Iglesias y Meriem El Yamri, por elegirnos para llevar a cabo este proyecto, habernos guiado con confianza en cada paso que dábamos y siempre sacar el lado bueno de las cosas aunque surgieran dificultades. Por trabajar a nuestro lado y confiar en nosotros hasta el último momento.

Muchas gracias.

Resumen

El aburrimiento es uno de los principales problemas de los entornos educativos del siglo XXI. La sociedad actual está hiperconectada y expuesta a una gran cantidad de estímulos diarios. Esta sobreestimulación cotidiana puede tener un efecto negativo en los estudiantes que se enfrentan a varias horas de clases online (especialmente en el nuevo contexto derivado del COVID 19) en las que experimentan escasez de estímulos, lo que se traduce en aburrimiento y, consecuentemente, en una pérdida de la atención frente al contenido académico.

Estudios pasados han demostrado que existen patrones faciales para detectar el interés y el aburrimiento. Los más perceptibles son el ensanchamiento de los párpados, la apertura de los labios y la dirección de la mirada. En cuanto a la postura, encontramos referencias al uso de la mano para sostener la cabeza, mantener una postura recostada u ocultar parte del rostro. Por otro lado, su contraparte, el interés, puede verse reflejado en algunos patrones como mantener la mirada con los ojos ampliamente abiertos, esbozar una sonrisa o mantener una postura erguida.

En este Trabajo de Fin de Grado se presenta el desarrollo, la arquitectura y la evaluación de una herramienta llamada VITAL EMO que busca: 1) identificar, a través de una cámara, el aburrimiento en los estudiantes, por medio del entrenamiento de una inteligencia artificial, y 2) alertar al profesor de esta situación, permitiéndole cambiar la metodología o estrategia para volver a captar la atención de los estudiantes.

La creación del dataset para entrenar a la IA se llevó a cabo con 12 personas. La participación tuvo lugar de manera telemática, donde cada participante visualizaba un vídeo mientras era grabado y respondía unas preguntas sobre el mismo.

Los resultados del estudio muestran que es posible detectar patrones faciales asociados al interés y al aburrimiento como pueden ser la sonrisa, los ojos entrecerrados, la oclusión del rostro o el no estar directamente frente a la cámara, a pesar del tamaño del dataset utilizado.

Estos resultados abren el camino a futuras investigaciones que, con dataset mayores, puedan mejorar la eficacia de las IAs que detecten el aburrimiento en los estudiantes online.

Palabras clave: Aburrimiento, emociones, educación, detector, inteligencia artificial, aprendizaje automático, redes convolucionales, aplicación de ordenador, visión por computador, vital emo.

Abstract

Boredom is one of the main problems of 21st-century educational environments. Today's society is hyperconnected and exposed to a large number of daily stimuli. This daily overstimulation can negatively affect students facing several hours of online classes (especially in the new context derived from COVID-19). Students experience a shortage of stimuli in these classes, resulting in boredom and a shorter attention span.

Past studies have shown that specific facial patterns help detect lack of engagement and boredom. The most noticeable are the widening of the eyelids, the opening of the lips, and the gaze direction. As for posture, there are references for boredom indicators for using the hand to support the head, maintaining a reclining posture, or hiding part of the face. Its counterpart, engagement, can be reflected in patterns such as maintaining a wide-eyed gaze, smiling, or maintaining an upright posture.

This end-of-degree project presents the development, architecture, and evaluation of a tool called VITAL EMO. Through artificial intelligence (trained with facial patterns), this tool seeks 1) to detect, using a camera, students' boredom by training an AI and 2) to alert teachers of this situation, allowing them to change their methodology or strategy to reengage the students over the go.

The creation of the dataset used to train the AI was carried out with 12 people. The participants joined telematically; each participant was recorded as they watched a video. Despite using a small dataset, the study results show that it is possible to detect facial patterns associated with engagement and boredom, such as smiling, squinting, facial occlusion, or avoiding being directly in front of the camera.

These results pave the way for future research to increase the effectiveness of AIs specialized in detecting online students boredom.

Keywords: Boredom, emotions, education, detector, artificial intelligence, machine learning, convolutional networks, computer application, computer vision, vital emo.

Índice general

	Página
1. Planteamiento del problema y justificación de la investigación	7
1.1. Problema detectado	7
1.2. Soluciones planteadas por otros	7
1.3. Aplicación exitosa del reconocimiento de emociones	8
1.4. Propuesta planteada	8
1.5. Estructura de la memoria	8
1. Approach to the problem and resolution	10
1.1. Detected problem	10
1.2. Proposed solutions by other researchers	10
1.3. Successful application of emotion recognition	11
1.4. Proposed solution	11
1.5. Structure of the document	11
2. Fundamentación teórica y estado de la cuestión	13
2.1. El aburrimiento	13
2.1.1. Un enfoque de autorregulación para entender la tendencia al aburrimiento	13
2.1.2. Caracterizando la firma psicofisiológica del aburrimiento	14
2.1.3. Aburrimiento, atención sostenida y red de modo predeterminado	15
2.1.4. El cerebro aburrido: corteza insular y la red de modo predeterminado	16
2.1.5. Expresiones faciales que denotan interés o aburrimiento	18
2.2. Introducción a las emociones	18
2.2.1. Emociones “Positivas” y “Negativas”	18
2.2.2. Las emociones en el aprendizaje online	19
2.3. Detección	20
2.3.1. Detección analógica del aburrimiento	20
2.3.2. Detección digital de las emociones	24
2.3.3. Datasets	27
2.4. Reconocimiento facial	29
2.4.1. Herramientas de reconocimiento facial	30
2.4.2. Modelos usados en el reconocimiento facial	32
2.4.3. Problemas	35
2.5. Aplicaciones de la Inteligencia Artificial	36
2.5.1. Introducción a la Inteligencia Artificial	36
2.5.2. Aplicaciones en el ámbito educativo	45

2.5.3. Aplicaciones en el ámbito educativo	46
3. Objetivos	51
3.1. Objetivos y plan de trabajo	51
4. Metodología	54
4.1. Reacción directa al vídeo	54
4.2. Amazon MTurk	56
5. Arquitectura	60
6. Resultados	64
7. Limitaciones del estudio	69
8. Conclusiones	70
8.1. Futuras líneas de trabajo/actuación	71
8. Conclusions	73
9. Contribución	75
9.1. Diego Isar Muñoz	76
9.2. Álvaro David Ortiz Marchut	78
9.3. Daniel Padilla Rodríguez	80
9.4. Sofía Prieto Ibáñez	82
10. Bibliografía y enlaces de referencia	87
9.4.1. Capacidad humana de decisión y Supervisión	95
9.4.2. Robustez técnica y seguridad	98
9.4.3. Privacidad y Data Governance	103
9.4.4. Transparencia	105
9.4.5. Diversidad, equidad y no discriminación	107
9.4.6. Bienestar social y ambiental	110
9.4.7. Responsabilidad	112

Capítulo 1

Planteamiento del problema y justificación de la investigación

1.1. Problema detectado

En la actualidad el aburrimiento se ha convertido en uno de los estados más comunes que se experimentan diariamente. No es de extrañar ya que vivimos en una sociedad hiperconectada, necesitada y habituada a estímulos e interacciones constantes y que, en el caso de carecer de este nivel de estímulos, cae en un estado de aburrimiento (Communications, 2021).

Hasta el momento, muchos investigadores, como James Danckert y Colleen Merrifield, han tratado esta emoción desde diferentes perspectivas y acercamientos, sin embargo, al observarlo desde un punto de vista educativo, el número de investigaciones centradas en la importancia de todo el rango de emociones en el aprendizaje es escasa (Pekrun, 2005).

Si unimos este contexto a la nueva actualidad provocada por la aparición del Covid-19, en la que muchos centros educativos se han visto en la obligación de cerrar y modificar toda su actividad docente a un medio online, encontramos una gran carencia de información aplicable por parte de los pedagogos a la actividad docente, teniendo en cuenta el estado emocional de los alumnos, mediante el uso de plataformas online.

1.2. Soluciones planteadas por otros

Entre las soluciones propuestas hasta el momento (ver Capítulo 2), podemos ver planteamientos de aplicaciones que graban el comportamiento de los alumnos durante las lecciones impartidas, de las cuales, al procesar los datos y de manera diferida, se obtiene un resumen de las emociones de cada alumno (Zeng et al., 2020).

Sin embargo, estas aplicaciones no se encuentran en el mercado, además, ninguna de ellas nos proporciona una respuesta concisa de cómo proporcionar un soporte a los pedagogos, que les permita comprender de manera clara cuándo un alumno se está sintiendo aburrido y, por lo tanto, permita al docente modificar su método de impartir clase para paliar esa situación de aburrimiento.

1.3. Aplicación exitosa del reconocimiento de emociones

En nuestra investigación no hemos encontrado aplicaciones centradas en un ámbito tan concreto, como es la detección del aburrimiento en el área educativa, para obtener un feedback suficiente de su recorrido y éxito, por lo que nos centraremos en evaluar el éxito o el fracaso en una perspectiva más amplia como es la detección de emociones y su aplicación en un diferentes ámbitos.

No es sorprendente el hecho de que un tema tan complejo como el reconocimiento de emociones haya sido muy explorado para su uso en marketing, educación o incluso ocio. Algunos de estos ejemplos podemos encontrarlos en aplicaciones implementadas por Microsoft (2010) o el MIT (Matheson, 2017), teniendo estas dos grandes organizaciones multitud de proyectos en innovación.

Asimismo, en la actualidad existen una gran variedad de plataformas que ofrecen servicios de pago para implementar el reconocimiento de emociones, además de que empresas tan importantes como Apple han comprado pequeñas start-ups dedicadas a este sector para así fusionarlo con sus productos (ElBoletín, 2016).

Podemos sacar en claro que si entidades como éstas han invertido tiempo y dinero en el sector de reconocimiento de emociones, es lógico asumir que es un mercado fructífero y aún por explotar.

1.4. Propuesta planteada

La propuesta de este TFG consiste en una herramienta que proporcione una ayuda extra al docente a la hora de analizar las expresiones faciales de sus alumnos. De esta manera el pedagogo puede impartir la clase teniendo en cuenta el estado de aburrimiento de sus oyentes y si es necesario, modificar su método de exposición para volver a captar su atención.

Esta herramienta se ha planteado de manera que las mediciones de las expresiones faciales se obtengan de aplicaciones online, por lo que ayudará a los profesores a comprender el estado de sus alumnos, especialmente cuando estos se encuentren online.

Durante el desarrollo del TFG, se nos planteó la oportunidad de presentar este proyecto en el congreso internacional ICCE 2021(International Conference on Computers in Education) (ICCE, 2021). El artículo, *Vital Emo: The boredom detector with a machine learning perspective* (ver Anexo 2), fue presentado en la subconferencia *Artificial Intelligence in Education/Intelligent Tutoring Systems* (AIED/ITS, 2021).

1.5. Estructura de la memoria

La estructura de este trabajo estará dividida en los siguientes apartados.

Capítulo 2 **Fundamentación teórica y estado de la cuestión**, recoge todo el contexto en el que se encuentra en ese momento el reconocimiento de emociones, incluido el aburrimiento, tanto desde un punto de vista teórico como práctico.

En el capítulo 3 **Objetivos**, encontraremos definidas las metas principales y secundarias que pretendemos alcanzar con este trabajo.

En el capítulo 4 **Metodología**, se explica el experimento que hemos llevado a cabo y la metodología utilizada para su implementación.

En el capítulo 5 **Arquitectura**, observaremos la explicación de las tecnologías que utilizamos, junto con la arquitectura planteada.

En el capítulo 6 **Resultados**, encontraremos los resultados obtenidos de la IA implementada.

En el capítulo 7 **Limitaciones del estudio**, expondremos la discusión de los resultados obtenidos, además de enumerar las limitaciones encontradas durante el desarrollo del proyecto.

El capítulo 8 **Conclusiones**, recoge las resoluciones obtenidas a partir de los resultados alcanzados, junto con la exposición de las mejoras que deberían realizarse en un trabajo futuro.

En el capítulo 9 **Contribución**, encontraremos una descripción del trabajo realizado por cada uno de los integrantes del equipo.

En el capítulo 10 **Referencias bibliográficas**, encontraremos todas las referencias a trabajos citados.

Y para finalizar, los **Anexos**, se recoge la base ética que debe seguir la IA y la verificación de su cumplimiento además del artículo de investigación enviado a la conferencia internacional **ICCE (2021)**.

Chapter 1

Approach to the problem and resolution

1.1. Detected problem

Nowadays boredom has become one of the most common states experienced on a daily basis. It isn't surprising since we live in a hyperconnected society that needs and is used to constant stimuli and interactions. Without that level of stimuli we tend to fall into a state of boredom.

Until now, many researchers have dealt with this emotion from different perspectives and approaches, however, looking at it from an educational perspective, the number of investigations focused on the importance of the whole range of emotions in learning is scarce.

If we add this to the current situation caused by the appearance of Covid-19, in which many educational centers have been forced to close and modify all their teaching activity to an online format, we find a lack of information that pedagogues could apply to their teaching activity, taking into account the emotional state of the students when using online platforms.

1.2. Proposed solutions by other researchers

Among the solutions proposed so far, there are applications that record the behavior of students during their lessons, from which, by processing the data, a summary of the emotions of each student is generated and shown on a delayed basis.

Nevertheless, these applications aren't available on the market nowadays, and none of them bring us a concise answer on how to provide support to pedagogues, that allows them to clearly understand when a student is feeling bored and, therefore, help the teachers modify their teaching method.

1.3. Successful application of emotion recognition

In the conducted research we could not find any applications focused on such a specific field, as is the detection of boredom in the educational area, to obtain enough feedback of its path and success. Therefore, we will focus on evaluating the success or failure of a broader area: the detection of emotions and its application in a large number of fields.

It is not surprising that a topic as complex as emotion recognition has been widely explored in order to integrate it into marketing, education or even entertainment. Some of these examples can be found in applications implemented by Microsoft or MIT, with these two large organizations having multiple projects in innovation.

In addition, there are currently a wide variety of platforms that offer payment services to implement emotion recognition. These phenomena can be seen in companies as important as Apple, that bought small start-ups dedicated to this sector in order to merge them with their products.

It is clear that if entities such as these companies have invested time and money in the emotion recognition sector, it is logical to assume that it is a fruitful market that has yet to be exploited.

1.4. Proposed solution

The proposal of this end-of-degree project consists of a tool that provides an extra help to the teachers when analyzing the facial expressions of their students. In this way, the pedagogues can give the lecture taking into account the state of boredom of their listeners and, if necessary, modify their method of exposition to reengage the students over the go.

This tool has been designed in such a way that the measurements of facial expressions are obtained from online applications, so it will help the pedagogues understand the state of their students, not only in a face-to-face class but also while using the new online teaching platforms.

During the development of the end-of-degree project, we were given the opportunity to present this proposed solution at the international conference ICCE 2021 (International Conference on Computers in Education) (ICCE 2021 TOC, 2021). The paper, ***Vital Emo: The boredom detector with a machine learning perspective***(see Annex 2), was presented at the *Artificial Intelligence in Education/Intelligent Tutoring Systems* (AIED/ITS, 2021) subconference.

1.5. Structure of the document

The structure of this project will be divided into the following sections.

In chapter 2 **Theoretical foundation and state of the art**, we gather all the context in which emotion recognition, including boredom, is at that moment, both from a theoretical and practical perspective.

In chapter 3 **Objectives**, we will define the main and secondary goals that we intend to achieve with this project.

In chapter 4 **Methodology**, we explain the experiment we have carried out and the methodology used for its implementation.

In chapter 5 **Architecture**, we will see the explanation of the technologies we use, together with the proposed architecture.

In chapter 6 **Results**, we will find the results obtained from the implemented AI.

In chapter 7 **Limitations**, we will present the discussion of the results obtained, in addition to listing the limitations encountered during the development of the project.

Chapter 8 **Conclusions**, contains the resolutions from the obtained results, together with the presentation of the improvements that should be made in future work.

Chapter 9 **Contribution**, contains a description of the work carried out by each of the team members.

In chapter 10 **References**, we will find all the references to cited works.

And finally, in **Annexes**, we include the ethical basis that the AI should follow and the verification of its compliance, together with the research article submitted to the **ICCE (2021)** international conference.

Capítulo 2

Fundamentación teórica y estado de la cuestión

2.1. El aburrimiento

“El aburrimiento es una experiencia humana omnipresente que se puede describir mejor como una incapacidad para interactuar con el propio medio ambiente a pesar de la motivación para hacerlo.”

Aburrimiento, atención sostenida y red de modo predeterminado
Danckert and Merrifield (2018)

El aburrimiento es uno de los principales problemas a los que se enfrenta la sociedad del siglo XXI. Tanto para las empresas como para las familias, los smartphones nos han convertido en una sociedad hiperconectada (Communications, 2021), los estímulos en nuestras vidas son constantes, diariamente recibimos mensajes, notificaciones, noticias, anuncios de marketing... Todo ello en cuestión de segundos.

Esta disponibilidad inmediata de la información ha creado una sociedad inquieta en la que, si no se reciben constantes estímulos, tiende a aburrirse con relativa facilidad.

La sociedad ha cambiado, pero ¿el sistema educativo ha cambiado lo suficiente?

Los contenidos, cuanto más se avanza en el ámbito educativo, más densos y menos interactivos son y, por lo tanto, aunque exista motivación por parte de los estudiantes, estos contenidos se vuelven aburridos.

Este trabajo de fin de grado ahondará en dicho problema, desde una perspectiva más informática, y aplicaremos algoritmos de inteligencia artificial que permitan detectar el aburrimiento en las aulas.

2.1.1. Un enfoque de autorregulación para entender la tendencia al aburrimiento

Investigaciones recientes han concluido con la idea de que el aburrimiento es un factor importante a la hora de definir nuestro estado de ánimo (Struk et al., 2016). Esto pue-

de llegar a afectar hasta a las actividades que hacemos o incluso en las decisiones que tomamos día a día.

Se ha demostrado que los individuos que son más propensos al aburrimiento tienen más riesgo de no sentirse retados, sentir una falta de motivación, no encontrar el sentido a ciertas cosas que se realizan en la vida cotidiana, etc.

La tendencia al aburrimiento se ha asociado como un rasgo característico de los individuos que tienen una pobre autorregulación, como podría ser personas con problemas de ludopatía, abusos de sustancia o atracones de comida.

La medida más común para medir el aburrimiento, la Boredom Proneness Scale o BPS (Farmer and Sundberg, 1986) o Escala de propensión al aburrimiento, define el aburrimiento como una construcción multifactorial (Vodanovich et al., 2005) sugieren que el aburrimiento está dividido en dos factores dependiendo si el origen está generado por estimulación exterior o interior.

- **Externo:** Debido a problemas para satisfacer la necesidad por los retos, el excitamiento o los cambios. Caracterizado por la insensibilidad ante los errores y puede ser un indicativo de un posible problema de déficit de atención o un desorden de hiperactividad.
- **Interno:** Caracterizado por asociarse con problemas para generar un interés en la vida diaria o sentir un compromiso por lo que se hace. Se vincula con problemas de habilidad para identificar y procesar sentimientos. Por ejemplo: la confusión que hemos tenido todos al tirar la cuchara a la basura y el yogur al fregadero.

Las personas con un alto nivel de autocontrol son menos propensas al aburrimiento independientemente de la fuente. Un aspecto crítico de la autorregulación es la flexibilidad ante el cambio. Los individuos con mayor flexibilidad para seleccionar cómo afrontar un problema o tomar decisiones ante un obstáculo suelen tener menos dificultades a la hora de encontrar motivación.

2.1.2. Caracterizando la firma psicofisiológica del aburrimiento

Investigadores como Mikulas and Vodanovich (1993) coinciden al afirmar que el aburrimiento es un estado poco placentero provocado por algo monótono.

Un área en la que entran en conflicto es la relacionada con la excitación que caracteriza al aburrimiento. Algunos investigadores consideraban que el aburrimiento debía asociarse con un aumento del nivel de excitación, como propuso Berlyne (1960), mientras que otros entendían el aburrimiento como un estado asociado a una bajada del nivel de excitación, como plantearon Mikulas y Vodanovich.

Tras realizar un experimento con personas a las que se les monitorizó el ritmo cardíaco, conductividad de la piel y nivel de cortisol; se logró asociar al aburrimiento con una disminución de conductividad en la piel y un aumento del ritmo cardíaco. Este resultado resultó muy chocante ya que asocia el aburrimiento a su vez con una subida y una bajada de la excitación. A este fenómeno se le denomina *directional fractionation* (fraccionamiento direccional) (Lacey, 1959) y provoca que la medición del ritmo cardíaco como única variable no sea del todo fiable.

Existen diferentes tipos de personas propensas al aburrimiento para las cuales la media del ritmo cardiaco es mayor y experimentan diferentes variaciones del aburrimiento. Además, hay personas que necesitan una gran estimulación interna, experimentan un aburrimiento 'apático' y tienden a sufrir pérdidas de atención diaria.

A diferencia de otros individuos propensos al aburrimiento que informan de una gran necesidad de estimulación externa, experimentan un aburrimiento 'agitado' y en su caso no tienden a sufrir estos lapsos diarios, pero les cuesta mantener la atención centrada en otras tareas. (Carriere et al., 2008; Cheyne et al., 2006; Malkovsky et al., 2012)

2.1.3. Aburrimiento, atención sostenida y red de modo predeterminado

El aburrimiento representa un estado de desinterés o desconexión, por ello, se puede esperar que existan varias áreas involucradas en esta distracción o falta de atención.

A estas regiones se las conoce como Default Mode Network (DMN - Redes en modo predeterminado), y se activan con el pensamiento interno, como por ejemplo, pensar en uno mismo, reflexionar en el pasado o imaginar el futuro (Andrews-Hanna, 2012; Buckner et al., 2008; Gusnard et al., 2001; Mason et al., 2007; Raichle et al., 2001).

Cuando una persona está involucrada en una tarea, la actividad central del cerebro se incrementa, mientras que la actividad de la DMN decrece (Greicius et al., 2003; Mason et al., 2007).

En el año 2009, Christoff y unos compañeros usaron un experimento para medir la distracción con un escáner. Dicho experimento trataba de hacer una tarea de atención continua (Christoff et al., 2009). Los resultados demostraron que hay una correlación entre la actividad de la DMN y la falta de atención.

Al investigar esto, Danckert and Merrifield (2018) encontraron una correlación entre la activación de las regiones posteriores del cerebro a través de cuatro acciones: ver un vídeo de inducción al aburrimiento, hacer una tarea de atención continua, ver un vídeo de inducción al interés y ver un vídeo de descanso. Todas las acciones tenían una duración de 8 minutos.

Para el experimento se usaron 13 adultos, de los cuales solo se procesaron los datos de 10 participantes.

Antes del experimento, se tenía que rellenar una encuesta en la que los participantes debían puntuar su estado de aburrimiento. Este servía de base para medir los resultados del experimento.

Las acciones detalladas son las siguientes:

- El vídeo de inducción al aburrimiento consistía en ver a dos hombres tendiendo la ropa.
- El vídeo de inducción al interés consistía en ver un documental de la BBC sobre el planeta tierra, con varios estímulos visuales.
- La tarea de atención continua consistía en ver una pantalla con aproximadamente 250 puntos blancos y, cada 2 o más segundos, aparecía o desaparecía un punto;

y, posteriormente, se tenía que pulsar un botón que servía para contabilizar las estrellas. Además se pedía localizar el sitio aproximado en el que se había producido el cambio.

- La tarea de descanso trataba de mirar una cruz negra centrada, con el fondo blanco.

El aparato usado para las acciones son unas gafas binoculares controladas por un ordenador.

Los datos que obtuvieron fueron preprocesados y analizados, cabe destacar que también tomaron muestras del nivel de oxígeno y la huella dactilar.

Los resultados de la encuesta mostraron que los participantes se aburririeron más en el vídeo de inducción al aburrimiento, la tarea de atención continua y la tarea de descanso, respecto a antes del experimento. En contraposición, la tarea del vídeo de inducción al interés disminuyó el estado del aburrimiento.

Finalmente, los resultados del escáner mostraron actividad cerebral en una zona concreta, siendo el vídeo de inducción al aburrimiento, la tarea de atención continua y la tarea de descanso, zonas que activan la DMN. En contraposición, la tarea del vídeo de inducción activa de forma inversamente correlacionada la misma zona.

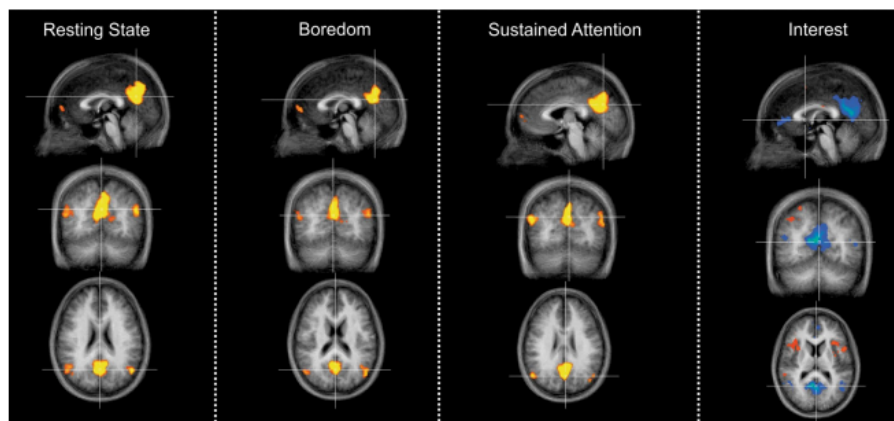


Figura 2.1: Escáner de actividad cerebral (Resting State, Boredom, Sustained Attention, Interest).

Fuente: Danckert and Merrifield (2018)

2.1.4. El cerebro aburrido: corteza insular y la red de modo pre-determinado

Un nuevo equipo de investigación compuesto por James Danckert y Julia Isacescu volvieron a replicar el experimento la correlación con el DMN y el aburrimiento (Danckert and Isacescu, 2017).

Se realizó el mismo experimento antes mencionado, esta vez fueron 15 participantes y solo se usaron datos de 13 de ellos.

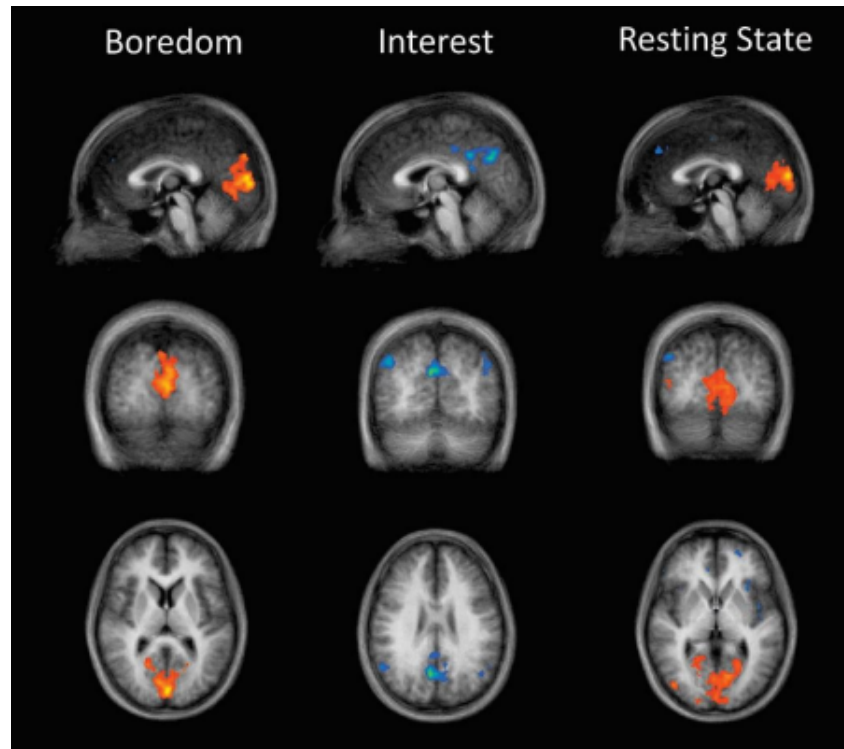


Figura 2.2: Escáner de actividad cerebral (Boredom, Interest, Resting State).

Fuente: Danckert and Isacescu (2017)

A diferencia del anterior estudio, esta vez sí encontraron actividad en la corteza insular. Esta actividad está relacionada con el estado de entretenimiento, y la diferencia de patrones entre los distintos estados y la actividad de la ínsula son indicativo de variación de niveles de desempeño.

En otros estudios se piensa que la corteza insular es esencial en el cambio cognitivo exterior e interior (Gao and Lin, 2012), y en el experimento se observa que durante las partes de inducción al aburrimiento y las de descanso tiene una menor actividad, mientras que en los vídeos interesantes la actividad es mayor.

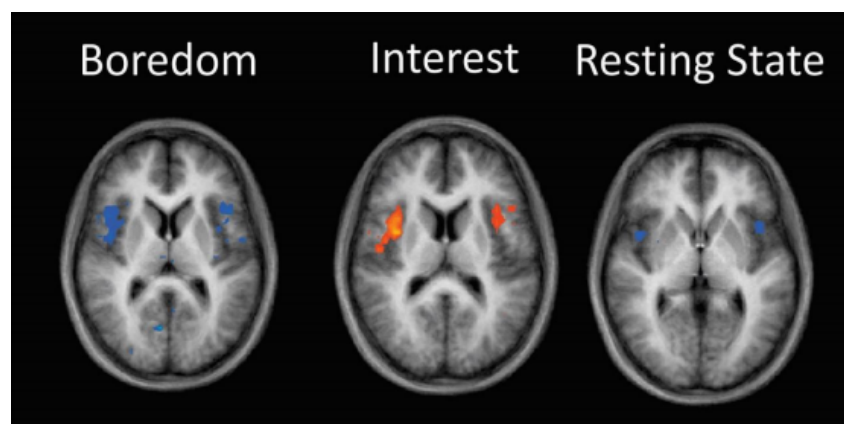


Figura 2.3: Escáner de actividad cerebral, perspectiva cenital.

Fuente: Danckert and Isacescu (2017)

2.1.5. Expresiones faciales que denotan interés o aburrimiento

Encontramos estudios que demuestran la existencia de patrones faciales para detectar si una persona se encuentra interesada o aburrida. Silvia (2001) mostró que los más perceptibles eran el **ensanchamiento de los párpados**, la **apertura de los labios** y la **dirección de la mirada**.

Se observaron estos mismos resultados en una investigación llevada a cabo por Reeve and Nix (1997) en el que desarrollaron un experimento para detectar los patrones faciales que caracterizan el interés. Para ello contaron con 60 personas, a las cuales se les hizo resolver puzzles mientras eran grabados. Obteniendo los patrones característicos y validando así las investigaciones que existían hasta la fecha.

2.2. Introducción a las emociones

2.2.1. Emociones “Positivas” y “Negativas”

Investigación en la que se intenta reestructurar la noción ya construida de valencia en las emociones así como el concepto de emociones opuestas o contrarias (Solomon and Stone, 2002). Por ejemplo, cuando pensamos en emociones opuestas, quizá pensemos en el amor y el odio. Sin embargo, la idea principal es que cada emoción no tiene un solo opuesto, si no que tiene un gran número dependiendo del contexto o la situación.

Algunas formas de clasificar las emociones junto con su opuesto pueden ser las siguientes:

Emoción positiva	Emoción negativa
Bueno	Malo
Placer	Dolor
Felicidad	Tristeza
Correcto	Equivocado
Virtud	Vicio
Aprobación	Desaprobación
Inquietante	Enervante
Sano	Insano
Calmado	Molesto

Cuadro 2.1: Emociones opuestas.
Fuente: Solomon and Stone (2002)

El filósofo holandés Spinoza sugiere que todas las “pasiones”, como él las categorizó, podrían ser catalogadas de una manera dentro de dos opciones. Por un lado están las que nos estimulan a la actividad, mientras que la otra opción es aquellas que nos drenan el interés por el mundo. La depresión podría ser un claro ejemplo de una emoción drenadora mientras que la alegría sería una emoción estimuladora. El problema de este pensamiento surge a la hora de clasificar muchas emociones que no producen una acción tan clara sobre el individuo.

Otra forma de clasificar las emociones está orientada a la idea de que las emociones negativas nos molestan mientras que las positivas no. Esta forma de pensar prioriza la

clasificación y el debate de las emociones malas mientras que las positivas quedan en un segundo plano siendo clasificadas como “las otras”. Fredrickson nos da tres explicaciones a esta priorización:

1. Un menor número de emociones positivas han sido diferenciadas y etiquetadas.
2. Los problemas psicológicos y de salud están producidos mayoritariamente por emociones negativas.
3. Los modelos de clasificación de emociones están centrados en emociones que producen una necesidad de acción como podrían ser el enfado o la molestia.

Nuevas tecnologías han permitido a los investigadores explorar el sistema inmunológico mientras varían los mecanismos psicológicos en los que influyen las emociones.

Cohen et al. (1995) descubrieron una relación entre los estados de ánimo negativos y la susceptibilidad a los virus respiratorios. Knapp et al. (1992) encontraron que tanto los estados positivos como los negativos tienen parecidos efectos en las funciones del sistema inmunológico produciendo un decremento de producción de linfocitos como respuesta a mitógenos.

Además de esto, surge el problema de no poder categorizar qué emociones producen felicidad y cuales son negativas. Esto es un concepto subjetivo que depende del individuo ya que está influenciado por la cultura, religión, contexto, historia, etc.

2.2.2. Las emociones en el aprendizaje online

La incorporación de las TIC como medio para el aprendizaje ha supuesto un reto para la adaptación de las anteriores metodologías didácticas.

Debido a esto, los investigadores Rebollo Catalán et al. (2008) han tratado de profundizar en la investigación que relaciona las emociones con el contexto educativo, pudiendo trasladarlo a estudios como la identificación de las emociones presentes en los procesos de aprendizaje (vergüenza, culpabilidad, orgullo, etc).

El psicólogo Pekrun (2005) habló de la importancia de las emociones en el aprendizaje y la falta de conocimiento que tenemos sobre ellas, especialmente en el ámbito online. Además, señala la falta de investigaciones en este campo, pudiendo encontrar solo referencias a estudios sobre la ansiedad relacionada con la evaluación y el rendimiento reflejado en exámenes y el estudio que relaciona las emociones con el éxito o fracaso académico.

Cabrera et al. (2006) resume algunos factores relacionados con el temprano abandono escolar, como son, la baja autoestima, incapacidad para sobrepasar dificultades, baja integración educativa y social. Otras investigaciones señalan que, además de estos, otros factores como la diferencia de género y el control de las emociones (Marchesi and Hernandez Gil, 2003; Rebollo-Catalán et al., 2006).

Con el objetivo de conocer y evaluar las emociones asociadas al aprendizaje, estos investigadores llevaron a cabo un estudio con estudiantes universitarios que atendían sus clases de manera semipresencial. Para ello pedían a los alumnos que evaluaran sus emociones tras las actividades de aprendizaje online y medían las emociones que los alumnos expresaban en los foros.

Al evaluar los resultados obtenidos observaron que las emociones relacionadas con el bienestar (orientación, alivio y optimismo) tenían un porcentaje más alto de aparición que las emociones asociadas con el malestar (tensión/preocupación y la desorientación).

Sacando como conclusión que las emociones como la orientación son de gran importancia para la formación de docentes en ámbitos de aprendizaje online, además de destacar la importancia de adquirir competencias emocionales para desarrollar un aprendizaje autónomo por parte de los estudiantes.

2.3. Detección

2.3.1. Detección analógica del aburrimiento

Herramientas utilizadas para la medición de emociones

Psicólogos como Robert Plutchik y Charles Osgood entre otros muchos han ideado diferentes herramientas que nos permiten acercarnos un poco más a los sentimientos de las personas.

- Semantic differential scale o prueba del diferencial semántico

Instrumento psicológico creado por Osgood et al. (1957) que permite calificar un producto, una comida o incluso una experiencia (Rosnow, 2004). Para ello, esta escala nos presenta una serie de adjetivos colocados de forma bipolar con un rango entre medias, permitiéndonos decidir en qué nivel nos encontramos.

Por ejemplo, al evaluar un partido político podremos ver una prueba de diferencial semántico parecida a esta.

Justo	1	2	3	4	5	Injusto
Imprudente	1	2	3	4	5	Prudente
Necio	1	2	3	4	5	Inteligente
Honesto	1	2	3	4	5	Deshonesto

Cuadro 2.2: Escala diferencial.

Osgood creó esta herramienta de manera que permitiera estudiar la estructura del significado de las emociones que provocan diferentes entidades, como la estancia en un hotel, la compra de un producto o la satisfacción de un empleado; concluyendo así que la estructura del significado tiene tres dimensiones.

- **Evaluación:** la relación entre bueno y malo.
- **Potencia:** la combinación entre poderoso y débil.
- **Actividad:** relación entre activo y pasivo.

Evaluar estos tres parámetros ayudaba al entendimiento de la emoción que estaba sintiendo la persona a la hora de evaluar la encuesta.

- Russell's circumflex model o círculo de emociones de Russell

Es un modelo en dos dimensiones en el que se dispone un amplio rango de emociones en un círculo donde los ejes están definidos, en horizontal por la valencia y en vertical por

la excitación (Russell, 1980). De este modo, dado un valor numérico de excitación y de valencia podemos identificar una emoción en concreto.

Este modelo ha sido muy criticado dado que hay emociones muy diferentes demasiado próximas. Un ejemplo podría ser la tristeza y la molestia.

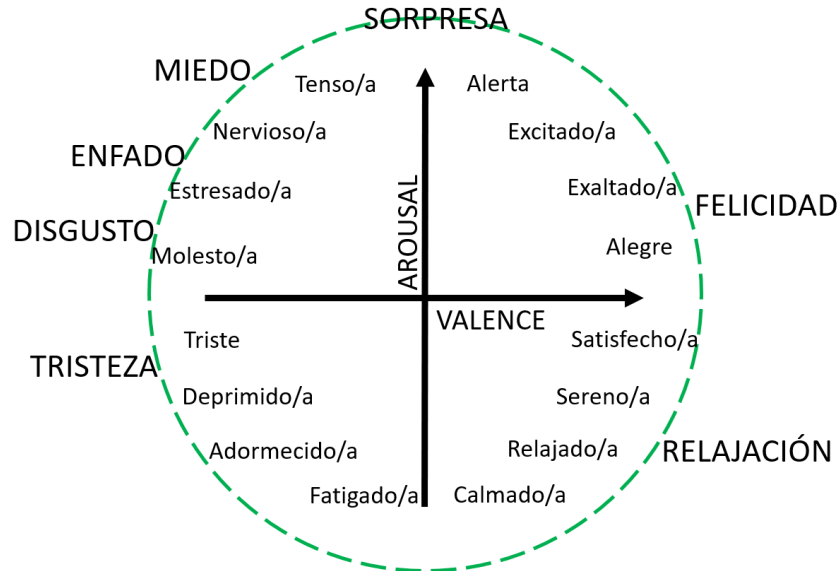


Figura 2.4: Círculo de emociones de Russell.

A pesar de las críticas hacia este modelo, se utiliza en un número relativamente alto de investigaciones y documentos.

- Self-Assessment Manikin (SAM) o maniquí de autoevaluación

Es un cuestionario pictográfico desarrollado por Bradley and Lang (1994) para la medición de las emociones.

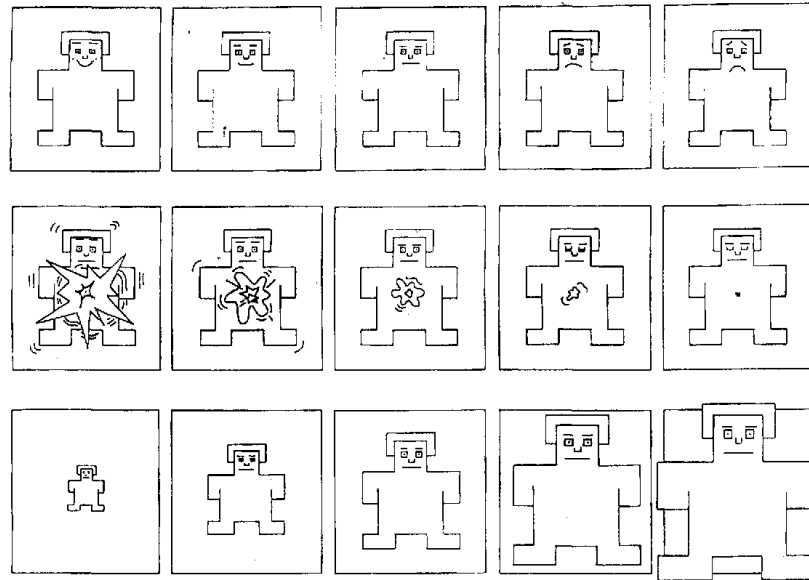


Figura 2.5: Cuestionario pictográfico de Bradley y Lang.
Fuente: Bradley and Lang (1994)

Se basa en imágenes sencillas sin texto, por lo que puede ser utilizado por cualquiera sin la necesidad de traducción. El encuestado debe elegir por cada una de las filas el muñeco que representa mejor la emoción que esté sintiendo.

Representando cada una de las filas:

- **Valencia:** la medición en relación a lo positivo o negativo
- **Excitación:** la evaluación del nivel de excitación, alta o baja
- **Dominancia:** la percepción de cuán empoderado se siente el sujeto

Estos tres aspectos a su vez caracterizan a las emociones, lo que permitía que con este cuestionario y el círculo de Russell, el reconocimiento de emociones pasara a ser algo más sencillo de entender y clasificar.

- Plutchik's wheel of emotions o rueda de las emociones de Plutchik

Es un modelo, creado por Plutchik (1991), el cual se organizaba bajo la existencia de 8 emociones básicas (felicidad, confianza, miedo, sorpresa, tristeza, anticipación, enfado y asco).

La rueda coloca los colores de manera que las emociones opuestas se encuentren enfrente y se localice las emociones compuestas entre los pétalos de la rueda. Si nos fijamos en cualquiera de los pétalos de dicha rueda observaremos como la intensidad de color va disminuyendo cuanto más nos alejamos del centro, haciendo referencia a como las emociones exteriores son menos intensas en comparación a las interiores.

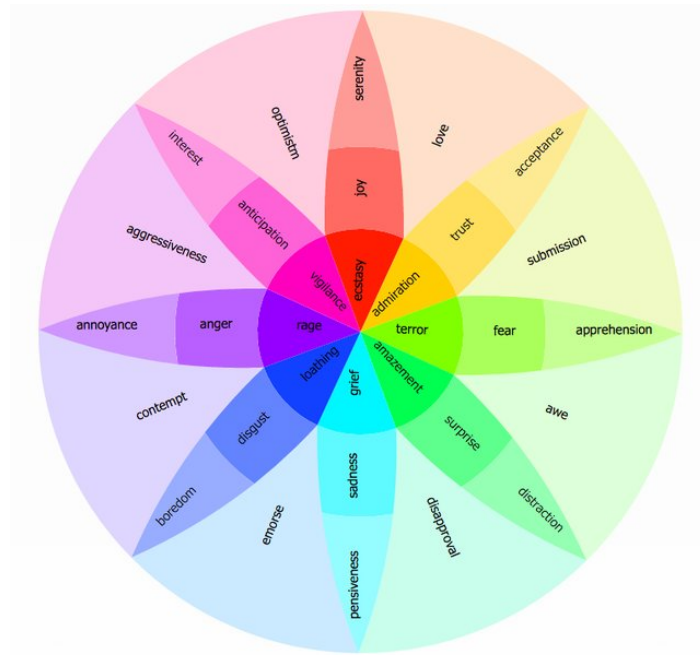


Figura 2.6: Rueda de las emociones de Plutchik.
Fuente: Nielek et al. (2017)

Esta rueda nos permite brindar algo de claridad sobre la identificación de emociones gracias a su amplio rango de emociones identificadas entre las cuales el encuestado podrá elegir.

- Boredom Proneness Scale (BPM) o escala de propensión al aburrimiento

Este test fue creado por Farmer and Sundberg (1986) con el objetivo de analizar la propensión de una persona al aburrimiento. Para conseguir esto analiza la estimulación externa, la percepción del tiempo, las limitaciones, las respuestas afectivas y la resistencia de la persona para mantenerse enfocado.

Está compuesto de 28 preguntas de verdadero o falso y puede adaptarse a personas de todas las edades.

- Beck Depression Inventory-II (BDI-II) o inventario de depresión de Beck

Cuestionario creado por el psiquiatra Beck et al. (1996), permite establecer el nivel de severidad de la depresión que sufre una persona. Está compuesto por 21 preguntas de respuesta múltiple que abarcan apartados como la irritabilidad, culpa y los síntomas físicos asociados a la depresión.

El resultado obtenido en dicho cuestionario está dividido en varios niveles:

- De 0 a 13 refleja niveles mínimos de depresión.
- De 14 a 19 refleja niveles medios de depresión.
- De 20 a 28 refleja niveles moderados de depresión.
- De 29 a 63 refleja niveles severos de depresión.

Detección del aburrimiento mediante el lenguaje corporal

El psicólogo Mehrabian et al. (1971) creó la regla del 7 % - 38 % - 55 %, la cual habla de la importancia de los diferentes canales por los que recibimos un mensaje, siendo estos las palabras, el tono de voz y la expresión facial y corporal. A estos canales les asignó una respectiva importancia de 7 %, 38 % y 55 %; haciendo referencia así a la gran importancia que tiene la comunicación no verbal a la hora de expresarnos.

Teniendo en cuenta una recopilación de vídeos grabados en conferencias, los cuales forman parte de la investigación internacional del AMI, se observaron una serie de signos los cuales denotaban aburrimiento. Estos signos están recogidos por Kroes (2007) en su paper Detecting Boredom in Meetings y son los siguientes:

- **Postura:** Cuando hablamos de la posición que conservamos a la hora de estar sentados, mantener una postura recostada sobre la silla es síntoma de una baja excitación, que al unirlo a un ambiente poco estimulante y signos de insatisfacción, señalaría un indicio de aburrimiento. Su contraparte, estar erguido, indicaría alta excitación e interés.



Figura 2.7: Postura de aburrimiento y postura de interés.

- **Movimiento:** En el caso de encontrarse en un estado de aburrimiento, los movimientos de esa persona serán repentinos, mientras que una persona atenta mantendrá unos movimientos constantes.
- **Uso de las manos:** Las manos pueden revelar mucha información. Las personas en un estado de aburrimiento tienden a utilizar sus manos para sujetar su cabeza o tratar de ocultar su rostro. En el caso de estar atento la persona mantendrá las manos relajadas encima de la mesa y las utilizará para apoyar sus argumentos a la hora de hablar.
- **Ojos:** Mantener la mirada y tener los ojos bien abiertos es significativo de estar activo y atento. Sin embargo, un claro indicador de aburrimiento podríamos detectarlo cuando la persona tiende a mantener los ojos parcialmente cerrados y parpadea asiduamente.
- **Boca:** La boca es un conocido indicador de aburrimiento. Nos permite detectarlo con un simple bostezo o al producirse periodos en los que se mantiene cerrada y con una expresión neutra.

2.3.2. Detección digital de las emociones

Herramientas de reconocimiento

- Reconocimiento con aprendizaje profundo basado en la respiración

Con los avances de la inteligencia artificial se ha comenzado a medir las señales fisiológicas para analizar la conducta afectiva. Muchas de estas señales como los electrocardiogramas (ECG), electroencefalogramas (EEG), electromiogramas (EMG), la presión arterial (BVP), la respuesta galvánica de la piel (GSR), temperatura o patrones en la respiración.

En la detección de emociones estas señales tienen un gran potencial ya que son manifestaciones involuntarias del cuerpo humano.

En esta investigación dirigida por Zhang et al. (2017a) se utilizó una versión simplificada del círculo de Russell para ayudar a detectar emociones. Esta teoría indica que todas las emociones están divididas en un espacio circular de dos dimensiones donde los ejes son la excitación y la valencia. En la versión simplificada se resumen todas las emociones en cuatro: felicidad, relajación, enfado o tristeza.

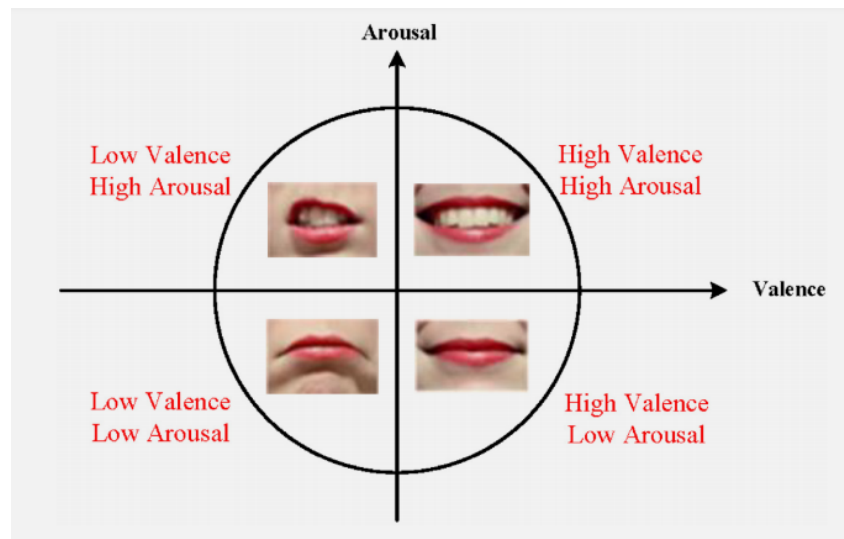


Figura 2.8: Versión resumida del círculo de emociones de Russell.

Fuente: Zhang et al. (2017a)

Para el experimento se utilizó el dataset de Dataset for Emotion Analysis using Physiological signals (DEAP) (Koelstra et al., 2012). Se obtuvo la reacción fisiológica de 32 participantes a la respuesta de 40 vídeos de música de un minuto cada uno.

Los datos respiratorios de estos individuos se tomaron con un sensor de respiración en forma de cinturón, que se coloca en el pecho, el cual mide la diferencia de tamaño del pecho durante la respiración.

Además de otros datos fisiológicos recogidos se le pidió a los participantes rellenar una encuesta en la que debían evaluar del 1 al 9 diferentes características sentidas en el momento como podría ser la excitación que sienten, la valencia, familiaridad, etc.

Se recogieron diferentes fragmentos de distintos tamaños (de 5 segundos a 60 segundos) de los datos respiratorios y se etiquetaron con la encuesta rellenada.

Estos datos se pasaban por un codificador de dos capas para hacer la entrada más pequeña

y manejable y la salida se utilizaba como entrada a dos regresiones lineales, una para detectar la excitación alta o baja y otra para detectar la valencia alta o baja.

La salida de cada regresión logística devolvía un resultado positivo o negativo que se utilizaba para identificar la emoción correspondiente en el círculo de Russell simplificado.

En los resultados se obtuvo una precisión del 73.06 % en la valencia y un 80.78 % en la excitación.

- Guante de mano para reconocimiento de gestos

Guante con distintos sensores para poder monitorizar la mano (Pławiak et al., 2016). Esto puede ser crucial a la hora de detectar ciertos gestos, posturas o movimientos que ayuden a identificar si la persona está aburrida.

Un posible modelo contiene sensores en cada dedo y giroscopios de modo que se está registrando en todo momento cualquier movimiento.



Figura 2.9: Guante con sensores.
Fuente: Pławiak et al. (2016)

- EEG para la medición de emociones

Un EEG o electroencefalograma se basa en una serie de electrodos colocados en posiciones claves del cuero cabelludo para la medición de la actividad cerebral.

La actividad cerebral se puede ver representada en forma de ondas, que dependiendo de su frecuencia y su localización se clasifican como ondas Gamma, Beta, Alfa, Theta y Delta.

- Gamma: 30 - 100+ Hz
- Beta: 12 - 30 Hz
- Alfa: 8 - 12 Hz
- Theta; 4 - 7 Hz
- Delta: 0 - 4 Hz

Siendo las ondas Gamma las que representan una mayor actividad cerebral, por ejemplo a la hora de memorizar, llegando hasta las ondas delta, que mostrarían un actividad casi nula.

Midiendo la actividad del sistema nervioso central podemos obtener el nivel de valencia y excitación que nos provocan las emociones, lo que nos permitiría diferenciarlas en positivas, negativas y neutras. Las ondas Beta y Gamma son buenas a la hora de reconocer la valencia de las emociones mientras que las ondas Theta ayudan al reconocimiento de las emociones excitantes.

En investigaciones como la de Lee and Hsieh (2014) podemos observar como las diferentes emociones positivas, negativas y neutras a las que son sometidos los participantes, proporcionan diferentes patrones en cuanto a valencia y excitación. Logrando así llegar a clasificarlas favorablemente utilizando un clasificador cuadrático de aprendizaje automático.

Basándose en ello, se intentó ampliar el rango de emociones utilizando el círculo de Russell, en el cual se clasifica las emociones teniendo en cuenta los parámetros de valencia y excitación.

Los investigadores Zamanian and Farsi (2018) identificaron las emociones de felicidad, tristeza, emoción y odio utilizando una SVM mejorada con un algoritmo genético llegando a obtener un 93,86 % de exactitud.

Múltiples proyectos se han llevado a cabo utilizando la tecnología de los EEG junto a otras, como por ejemplo en algoritmos para mejorar la recomendación, teniendo en cuenta las emociones, o utilizándolos para medir la atención en la conducción.

2.3.3. Datasets

Galvanic Skin Response (GSR)

Respuesta de conductancia de la piel frente a imágenes aversivas, excitantes o neutrales. Las imágenes usadas en este dataset proceden del Sistema Internacional de Imagen Afectiva o IAPS (Lang et al., 1997). Dichas imágenes están clasificadas como:

- Negativas
- Neutrales
- Positivas

Además tienen asignada una Valence (Valencia) y una Arousal (Excitación)

Sistema Internacional de Imagen Afectiva IAPS

Son imágenes ampliamente usadas en la investigación psicológica creadas por la Universidad de Florida (Lang et al., 1997), por lo que no suelen ser mostradas en ningún medio de comunicación o publicación. Por ello, su acceso es restringido y se necesita justificar su uso, siendo posible conseguirlas solo a través de una institución académica o de investigación, ambas sin ánimo de lucro.

Sonidos Digitalizados Afectivos Internacionales (IADS)

Son el equivalente de IAPS, pero en lugar de ser imágenes son sonidos (Stevenson and James, 2008).

Database for Emotion Analysis using Physiological Signals o DEAP

Análisis de los estados emocionales de los humanos a través de sus electroencefalografías (EEG) y señales fisiológicas de los participantes mediante grabaciones de vídeo (Koelstra et al., 2012). Esta base de datos se divide en varios datasets:

- Metadatos:
 - Lista de vídeos mostrados a los participantes
 - Puntuaciones asignadas por los participantes
 - Cuestionario rellenado por los participantes
 - Puntuaciones de los vídeos online
- Vídeo de las electroencefalografías de los participantes
- Vídeo de la cara de los participantes

Cabe destacar el uso de Valence (Valencia), Arousal (Excitamiento) y Dominance (Dominancia) en los datasets de las puntuaciones.

Facial Expression Recognition Challenge

Conjunto de imágenes publicadas en Kaggle utilizado como recurso en un concurso de Machine Learning (Carrier and Courville, 2013). El dataset está formado por más de 30.000 imágenes en escala de grises con unas dimensiones de 48 x 48 píxeles con su correspondiente emoción entre las que se diferencian:

- Enfado
- Desagrado
- Miedo
- Tristeza
- Sorpresa
- Neutral

2.4. Reconocimiento facial

Face Tracking captura de movimiento facial (Kim et al., 2008) es un concepto que aborda el seguimiento de una cara a través de cámaras. Consta de dos pasos (Schroff et al., 2015):

1. **Verificación / Detección de caras:** Se entrena a la red neuronal convolucional para que sea capaz de detectar una cara en una imagen. Para ello el dataset debe contener imágenes variadas, con objetos y caras debidamente etiquetadas y centradas. Posteriormente, se debe aplicar cualquiera de los siguientes algoritmos:
 - *Sliding Windows Detection* o detección de ventana deslizante (Chang et al., 2015): Al insertar una imagen cualquiera, se debe dividir la imagen en varias subimágenes y comprobar si detecta una cara, además el tamaño del recorte (Cuadro delimitador) debe variar porque una cara puede estar en varios cuadros delimitadores por lo que tiene un alto coste computacional.
 - *You Only Look Once* o YOLO (Redmon et al., 2016): Intentamos prever dónde está la cara, mediante predicciones que tengan altas probabilidades de acertar, y descartamos las que tengan un alto índice de intersección sobre la unión ($\text{Tamaño intersección} / \text{Tamaño total}$), para evitar resultados duplicados.
 - *Single Shot Multibox Detector* o SSD (Liu et al., 2016): Utiliza cuadros delimitadores de tamaño fijo pre-calculados, y al igual que en YOLO, se descartan las cajas que tengan un alto índice de intersección sobre la unión.
2. **Reconocimiento facial:** Una vez detectada la cara, podemos reconocer quién es dicha persona.
 - **Red Siamesa** (Guo et al., 2017): Codificamos las fotos y comparamos los vectores resultantes, si la diferencia entre los dos vectores es grande, son personas distintas, si es pequeña, son la misma persona. Para ello se usa:
 - **Triplet Loss Function** o función de pérdida de tripletes (Schroff et al., 2015). Codificamos una imagen ancla, una imagen positiva y una imagen

negativa, que sería la foto de la misma persona y la foto de una persona distinta respectivamente. Tras ello, calcula la distancia entre la imagen ancla y la imagen positiva, y la distancia entre la imagen ancla y la imagen negativa. Además, se debe establecer un margen α .

$$\max(f(A) - f(P)^2 - f(A) - f(N)^2 + \alpha, 0)$$

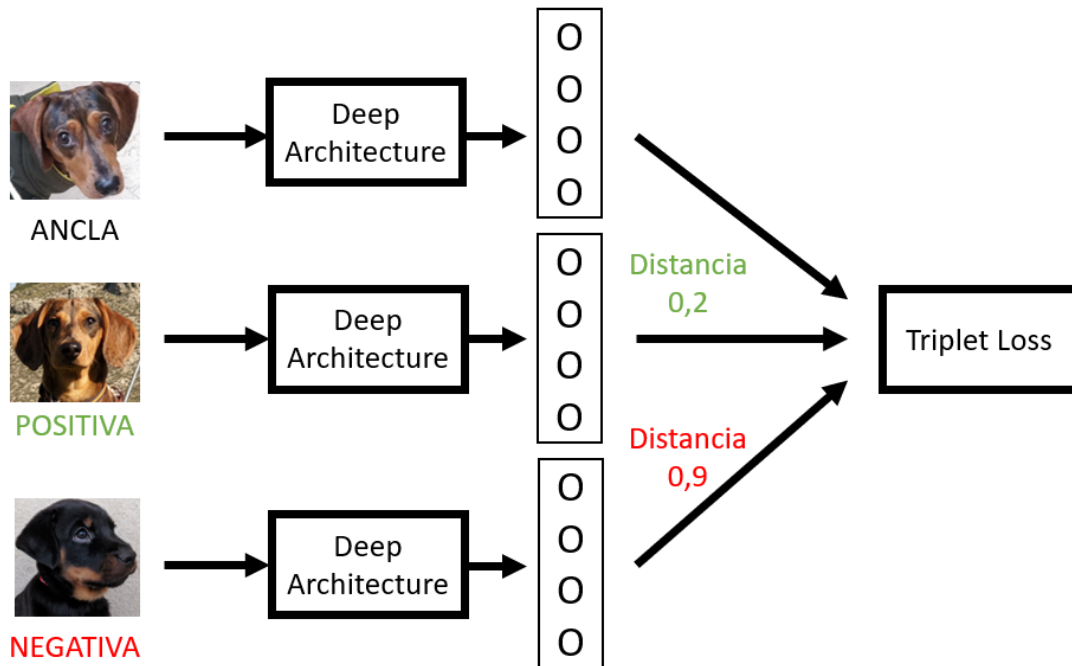


Figura 2.10: Ejemplo de funcionamiento de Triplet Lost.

- **Clasificación binaria** (Kumari and Kr., 2017): A diferencia de la Triplet Loss Function tenemos una red neuronal en la que codificamos dos caras, a ambas se les asigna una etiqueta que nos permitirá conocer si son la misma persona. En este caso, la predicción es una función sigmoide.

2.4.1. Herramientas de reconocimiento facial

OpenCV (Open Computer Vision)

Biblioteca libre de visión artificial desarrollada originalmente por Intel utilizada en el reconocimiento de objetos, detección de movimientos, reconstrucción 3D a partir de una imagen 2D, etc. (OpenCV, 1999)

Esta biblioteca incluye un clasificador ya entrenado para detectar caras (entre muchos otros clasificadores) que pueden llegar a ser útiles a la hora de eliminar toda la información innecesaria y quedarnos únicamente con el rostro de una persona para su posterior tratamiento y análisis.

Este proceso se consigue mediante el algoritmo Adaboost aplicando el framework de Viola-Jones especializado en detección de objetos.

Viola-Jones

Framework orientado en detección de objetos (especializado en detección de caras) que, transformando la imagen en escala de grises, divide la imagen en subregiones de diferentes tamaños y utiliza las llamadas Haar Features. (Wang, 2014)

Estas “Haar Features” nos permiten hacer comprobaciones sobre las sombras y contrastes de luz que se encuentran en las subimágenes que se han sacado de la imagen original. Por ejemplo, sabemos que la zona de los ojos es más oscura que su zona inferior (la mejilla), por lo tanto este filtro afirmará que esta subimagen contiene la parte central de la cara.

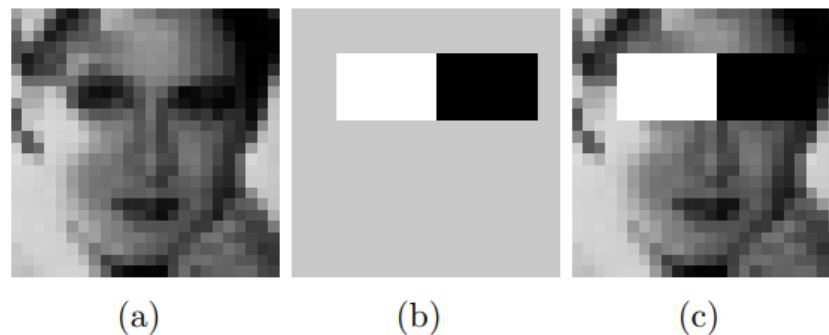


Figura 2.11: Ejemplo de funcionamiento de Viola Jones.

Fuente: Wang (2014)

Esta comprobación hace la función de un filtro en un algoritmo de Adaboost (M. Gopi Krishna et al., 2012), el cual está compuesto por miles de comprobaciones como estas. En caso de que un filtro no encuentre una porción de cara la imagen se descarta afirmando que no existe una cara en la imagen.

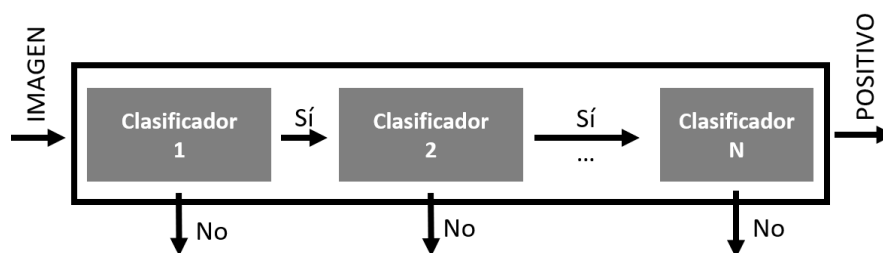


Figura 2.12: Ejemplo de funcionamiento de Adaboost.

Fuente: M. Gopi Krishna et al. (2012)

OpenFace

Biblioteca que, dada una imagen, nos permite reconocer rostros que se encuentren en ella mediante el uso de OpenCV y utilizando una red neuronal profunda modifica la imagen para centrarse de modo que parezca que la persona se encuentra de frente. (Amos et al., 2016)

Una vez realizado esto, el rostro se encuentra en una posición óptima para que el usuario aplique un algoritmo deseado de clustering para tratar la cara.

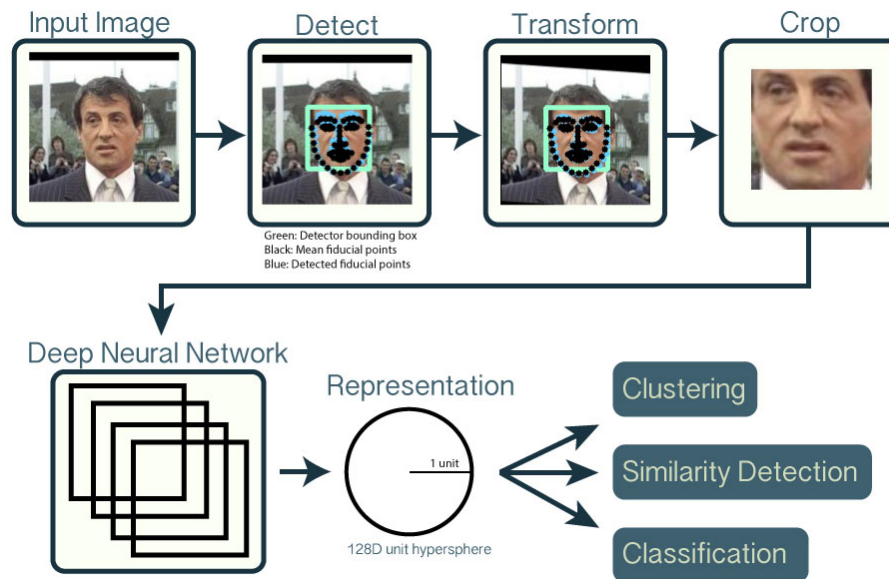


Figura 2.13: Ejemplo de funcionamiento de OpenFace.
Fuente: (Amos et al., 2016)

Reconocimiento facial basado en Local Binary Patterns

Metodología de reconocimiento de patrones y visión computacional (Ahonen et al., 2004). Se basa en dividir la imagen por sectores, se escoge un punto aleatorio y se mira a los vecinos comparándolo con el punto anterior. Dependiendo del resultado anterior, salen distintos patrones los cuales se pueden reconocer como puntos, esquinas, bordes, etc.

Actualmente está en desuso en pro de la metodología convolucional.

2.4.2. Modelos usados en el reconocimiento facial

Actualmente hay modelos probados que permiten una buena detección de objetos y caras, en la gran mayoría de papers que hacen referencia al face tracking hay 3 modelos que destacan:

- Modelo VGG16 Simonyan and Zisserman (2015)
- Modelo DeepFace Taigman et al. (2014)
- Modelo FaceNet Schroff et al. (2015)

Modelo VGG16

De los tres modelos es el más sencillo, ya que simplemente se debe reducir la imagen original a un tamaño de 224x224 en RGB y aplicar los siguientes pasos:

1. Reducir la imagen en un tamaño 224x224x3 (RGB).
2. Aplicar dos capas convolucionales de tipo SAME con la función de activación 'ReLU'.

3. Aplicar una capa Maxpooling que nos permitirá reducir el tamaño de la matriz a la mitad, esto se da porque $\text{stride} = 2$ y el tipo de convolución es VALID.
4. Aplicar dos capas convolucionales de tipo SAME con la función de activación 'ReLU'.
5. Aplicar una capa Maxpooling con $\text{stride} = 2$ y el tipo de convolución es VALID.
6. Aplicar tres capas convolucionales de tipo SAME con la función de activación 'ReLU'.
7. Aplicar una capa Maxpooling con $\text{stride} = 2$ y el tipo de convolución es VALID.
8. Aplicar tres capas convolucionales de tipo SAME con la función de activación 'ReLU'.
9. Aplicar una capa Maxpooling con $\text{stride} = 2$ y el tipo de convolución es VALID.
10. Aplicar tres capas convolucionales de tipo SAME con la función de activación 'ReLU'.
11. Aplicar una capa Maxpooling con $\text{stride} = 2$ y el tipo de convolución es VALID.
12. Aplicar dos capas Fully Connected con 4.096 neuronas y una con 1.000 neuronas.
13. Aplicar una capa Softmax.

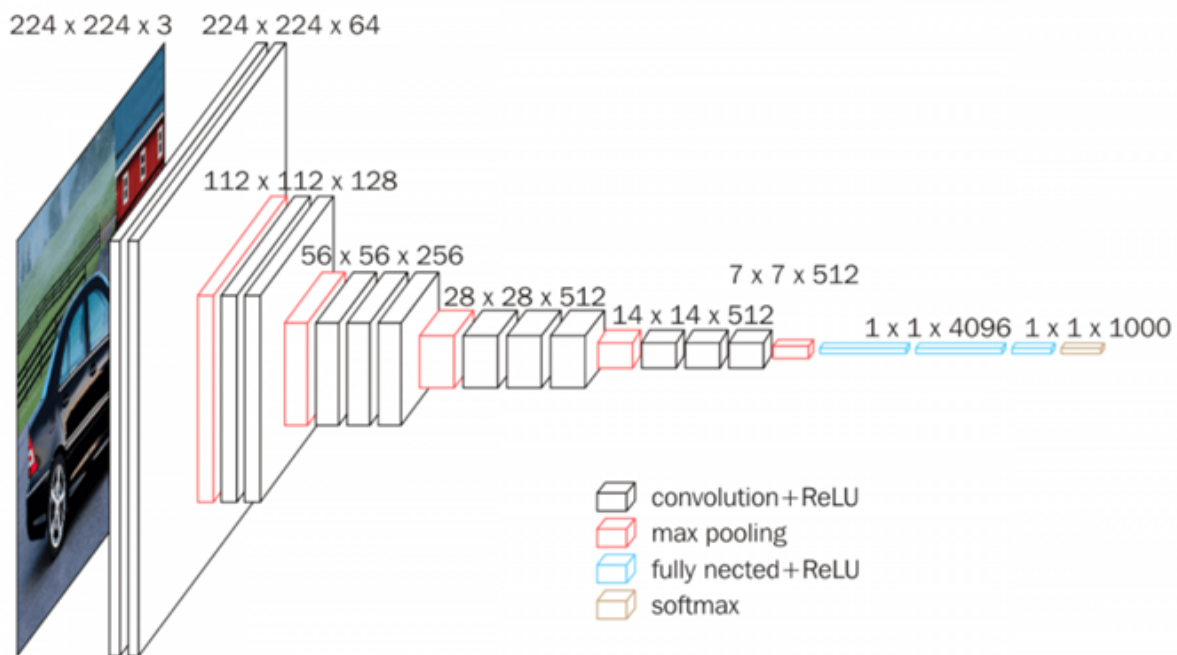


Figura 2.14: Modelo VGG16.
Fuente: Simonyan and Zisserman (2015)

Modelo DeepFace

Creado por Facebook, tiene una precisión aproximada del 97 %, desde un punto de vista técnico, es el modelo más difícil de implementar de los tres, y consta de 2 pasos, el primero se divide en:

1. Establecer unos landmarks o puntos de referencia iniciales, en este caso son seis.
2. Se hace un recorte 2D y se alinea la cara en el centro de la imagen.
3. Se generan 67 landmarks y se aplica la triangulación Delaunay.
4. Transformamos la imagen 2D a 3D y la rotamos de forma que queda completamente de frente.
5. Volvemos a transformar la imagen, de 3D a 2D.
6. Los landmarks nos sirven para ‘deformar’ la cara.
7. Recuperamos la imagen pero con la cara completamente de frente.
8. Residualmente podemos obtener un modelo 3D de la cara.

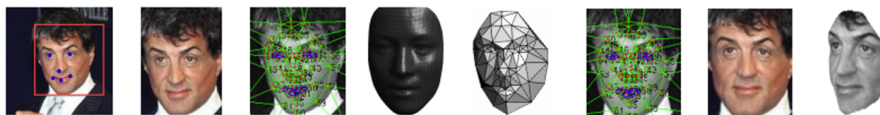


Figura 2.15: Modelo DeepFace Paso 1.

Fuente: Taigman et al. (2014)

El segundo paso es, tras tener la cara de frente:

1. Reducir la imagen a 152x152x3.
2. Aplicar una capa de convoluciones.
3. Aplicar una capa maxpool.
4. Aplicar una capa de convoluciones.
5. Aplicar tres capas Locally Connected, las podemos aplicar porque la cara, al estar transformada por el primer paso, siempre tendrá, por ejemplo, la nariz, en el centro de la imagen, o la boca en la zona central inferior.
6. Aplicar dos capas Fully Connected.

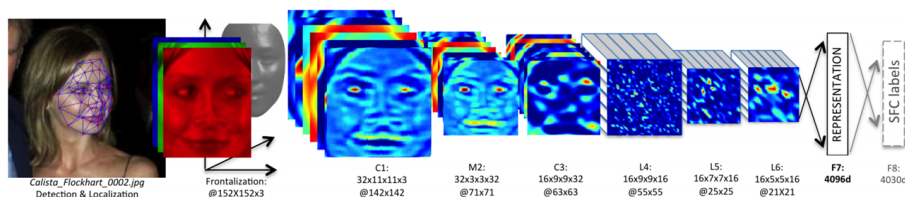


Figura 2.16: Modelo DeepFace Paso 2.

Fuente: Taigman et al. (2014)

Modelo FaceNet Creado por Google, tiene una precisión aproximada del 99'6 %. Este modelo permite tres funcionalidades:

- **Verificación:** ¿Es la misma persona?
- **Reconocimiento:** ¿Esta persona está dada de alta en la Base de Datos?
- **Clustering:** ¿Encontrar personas que se parezcan a X?

Este modelo basa su funcionamiento en la Triplet Loss Function (Schroff et al., 2015), explicado anteriormente.

2.4.3. Problemas

El problema de los anteriores modelos es que detectan la cara en cada imagen, por ello no están optimizados para utilizarse en un vídeo, siendo el coste computacional muy elevado. Además, se pierde la información entre frames, como por ejemplo, solo teniendo una imagen ¿cómo se distinguiría a una persona que está durmiendo de una persona que está parpadeando?

Seguimiento y reconocimiento facial consistente y preciso en vídeos

En una tesis presentada por Liu (2020), en la Universidad de Ohio, se propone una solución para los problemas del punto anterior, además de mejorar la consistencia en caras pequeñas, en ángulos complicados e inclusive borrosas. La solución consta de los siguientes pasos:

1. Se divide la entrada de vídeo en fotogramas.
2. Se usa una red neuronal convolucional, en este caso S3FD (Zhang et al., 2017a) que permite:
 - Recopilar estadísticas
 - Inferir la ubicación de la cara a través de una *Bounding Box*
3. Se reconoce la cara de la *Bounding Box* a través de la *Triplet Loss Function* (Schroff et al., 2015) o Función de Pérdida de Tripletes.
4. Guardamos la ubicación de la *Bounding Box* en una red neuronal recurrente, en este caso usa *Long Short-Term Memory* o LSTM (Hochreiter and Schmidhuber, 1997).
5. Se muestra la *Bounding Box* y la persona detectada en el frame.
6. Se pasa a la siguiente imagen y se aplica todo el proceso de detección.
 - Si la cara se detecta, se actualiza LSTM
 - Si la cara no se detecta, se usa la posición de la cara guardada en LSTM

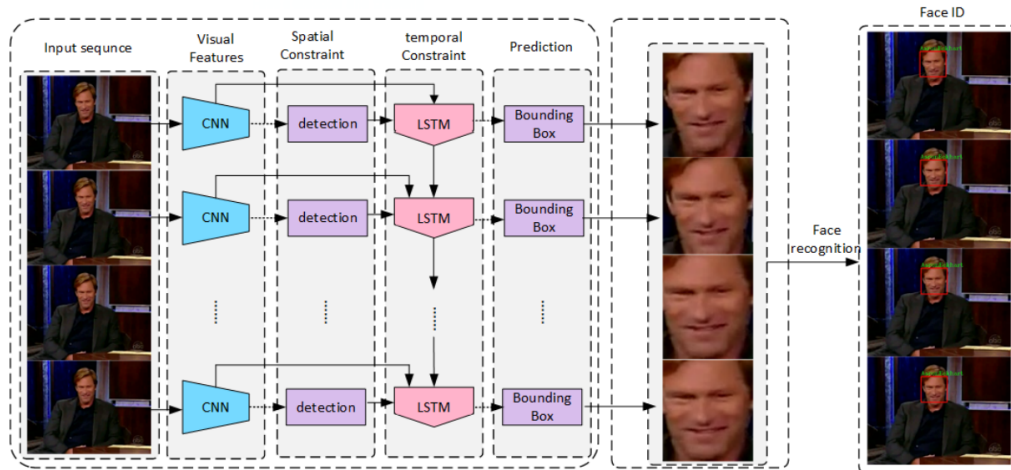


Figura 2.17: Arquitectura propuesta por la tesis.

Fuente: Liu (2020)

2.5. Aplicaciones de la Inteligencia Artificial

2.5.1. Introducción a la Inteligencia Artificial

La inteligencia artificial es la capacidad de los programas informáticos que, a través del uso de distintos algoritmos, permite emular las capacidades cognitivas de los seres humanos.

En 1950, Alan Turing crea “El test de Turing”, este test consiste en que una máquina debe engañar a un humano haciéndole creer que ella misma es un ser humano (Gonzalez Pacheco, 2019).

Posteriormente, en 1956, John McCarthy, matemático estadounidense, le da nombre a esta disciplina de la informática y, un año más tarde, Frank Rosenblat, un psicólogo estadounidense, crea la primera red neuronal (Alandete, 2011).

No fue hasta los años 90 que la Inteligencia Artificial tuvo un gran crecimiento, esto se debe a los avances tecnológicos de la época, como el aumento de la capacidad de procesamiento y el análisis de datos. Finalmente, en el año 1997, IBM lanzó Deep Blue, un ordenador que fue capaz de derrotar al, por entonces campeón del mundo de ajedrez, Gari Kaspárov (de la Torre, 2019).

Machine Learning

Dentro de la Inteligencia Artificial, existe un subconjunto de técnicas denominadas Machine Learning o Aprendizaje Automático (Schmitt, 2020). Según Tom Mitchell, pionero en este sector:

“El Aprendizaje Automático es el estudio de algoritmos informáticos que permite a los programas mejorar de forma automática a través la experiencia”

Tom Mitchell

Este aprendizaje a su vez se divide en tres paradigmas (Recuero de los Santos, 2017):

- **Aprendizaje supervisado:** Intenta predecir la respuesta correcta para un único ejemplo.

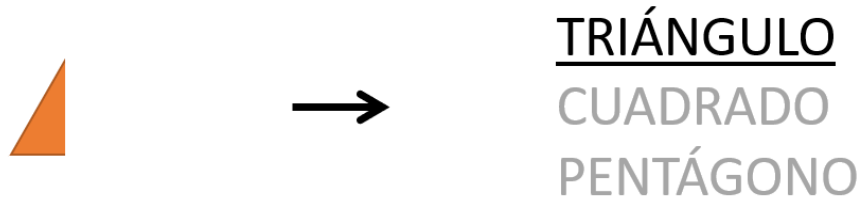


Figura 2.18: Ejemplo de aprendizaje supervisado.

- **Aprendizaje no supervisado:** Busca agrupar ejemplos dadas sus características.

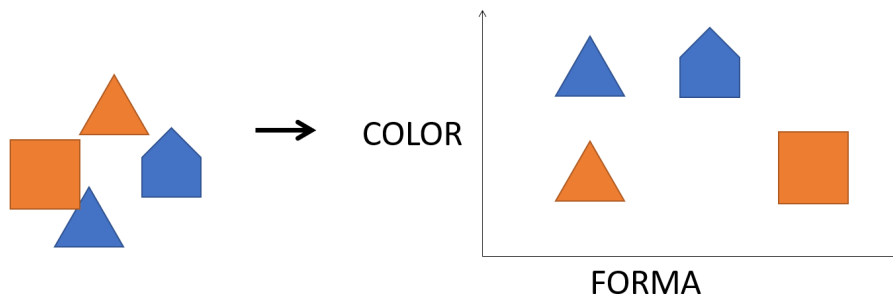


Figura 2.19: Ejemplo de aprendizaje no supervisado.

- **Aprendizaje por refuerzo:** Un agente está programado para maximizar la recompensa acumulada por cada acción correcta en un entorno.

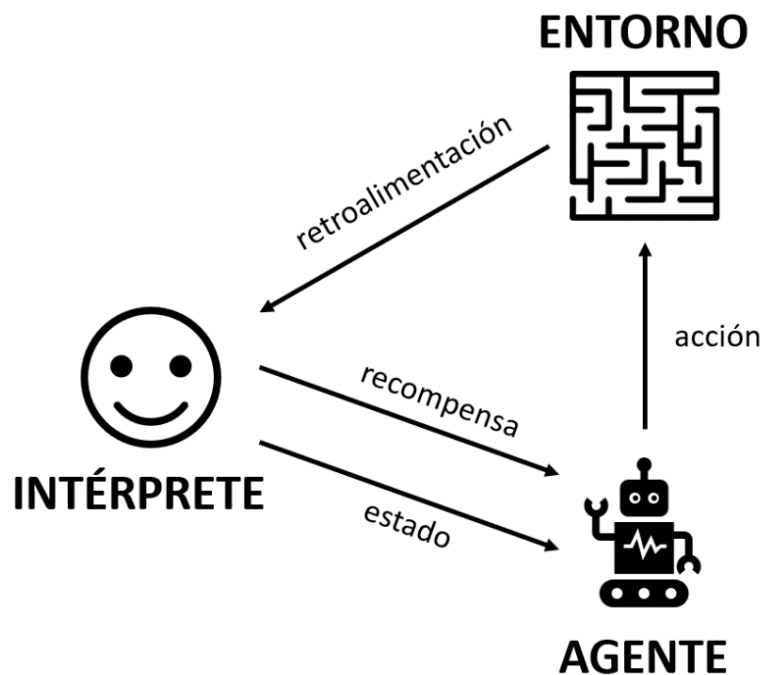


Figura 2.20: Ejemplo de aprendizaje por refuerzo.

Deep Learning

Finalmente, dentro del machine learning existe un conjunto de técnicas y algoritmos que imitan el aprendizaje de los seres humanos respecto a una determinada tarea, a esto se le denomina deep learning o aprendizaje profundo (Oppermann, 2019). Este tipo de aprendizaje se basa en una composición de Redes Neuronales.

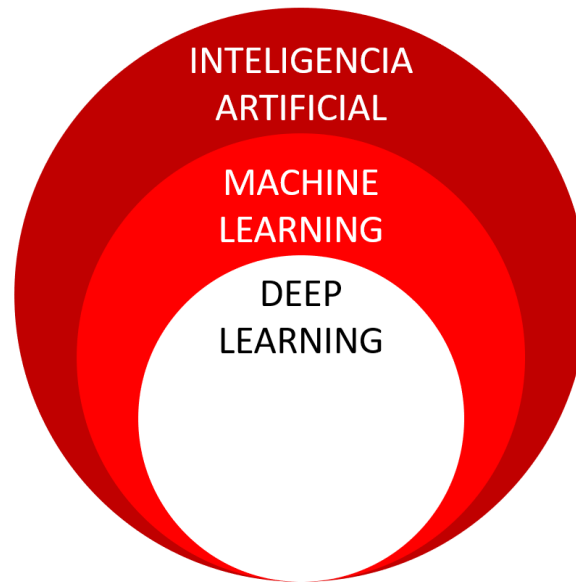


Figura 2.21: Jerarquía de conceptos respecto a la Inteligencia Artificial.

Redes Neuronales

Las redes neuronales son un modelo que está compuesto por unidades.

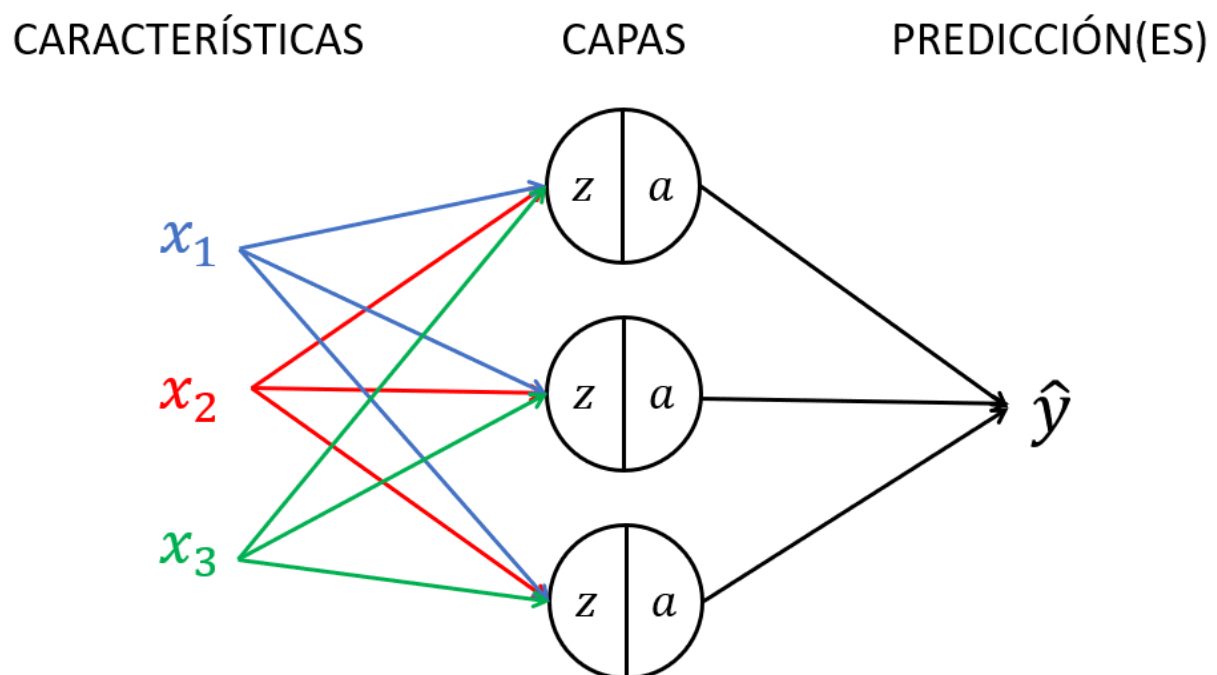


Figura 2.22: Ilustración del funcionamiento conceptual de una red neuronal.

El funcionamiento de estas unidades se divide en dos pasos:

- **Forward Propagation**

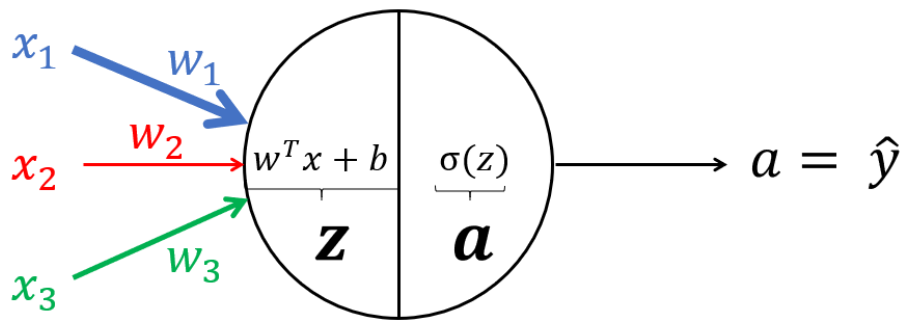


Figura 2.23: Ilustración del funcionamiento lógico de una neurona artificial.

Permite calcular el error a través de una entrada de características ‘ X ’, se realiza una operación matemática, en este caso la multiplicación del peso ‘ w ’ y la característica ‘ x ’ más el bias ‘ b ’.

- **w:** valor de los enlaces, incrementan o disminuyen su valor para minimizar el coste de la función de pérdida (En el siguiente paso).
- **x:** valor de la característica o de la anterior neurona.
- **b:** valor que nos permitirá controlar el sesgo que tiene cada neurona.

Posteriormente, sobre el resultado anterior ‘ z ’, se utilizan funciones de activación, que evitan que las capas intermedias sean funciones lineales. Hay distintas funciones de activación, entre ellas se encuentran:

- **Función sigmoide:** Se usa normalmente para hacer una clasificación binaria.

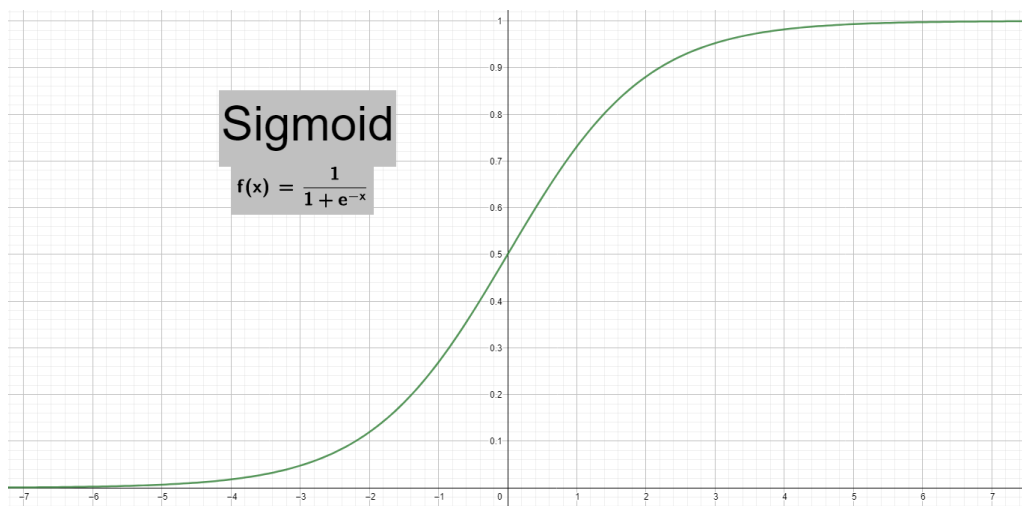


Figura 2.24: Función sigmoide.

- **Tangente hiperbólica:** Ofrece mejores resultados que la sigmoide.

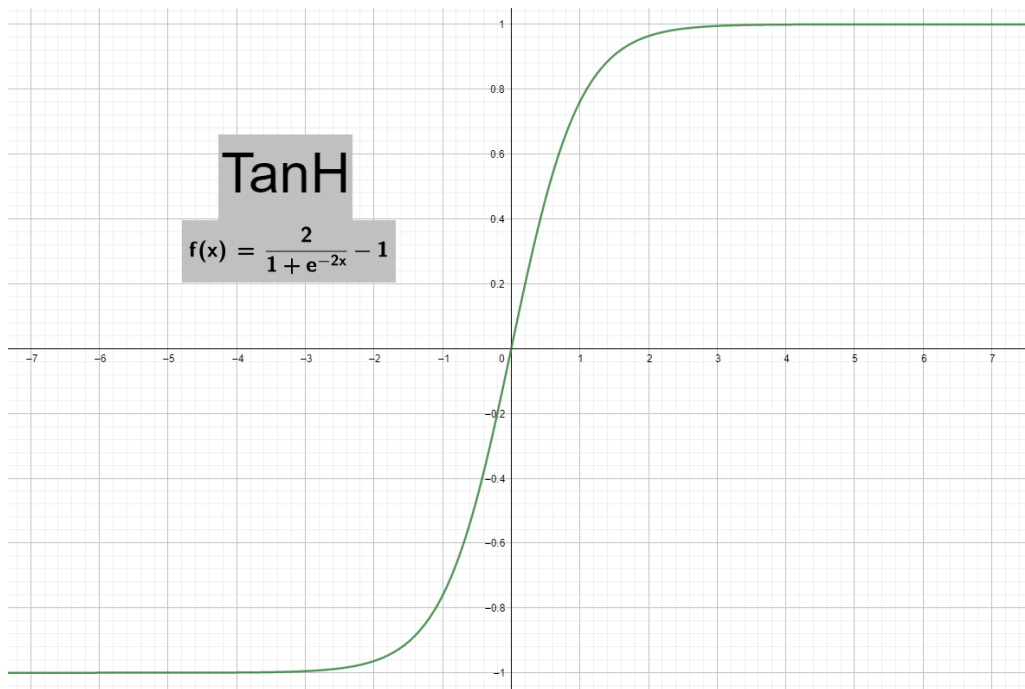


Figura 2.25: Función de tangente hiperbólica.

- **Rectified Linear Unit (ReLU):** Aprende más rápido que la sigmoide y la tanh, puede ser menos precisa.

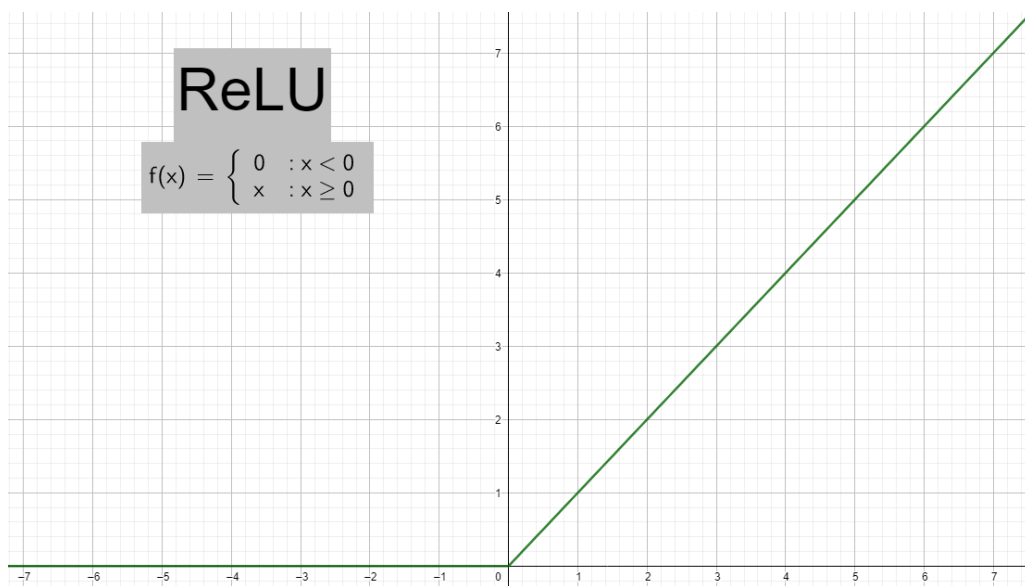


Figura 2.26: Función ReLU.

- **Softmax:** Es una función de activación que suele ubicar al final de la red neuronal, dado que proporciona una salida comprendida en el rango $[0, 1]$, siendo el sumatorio de todas las salidas $= 1$.

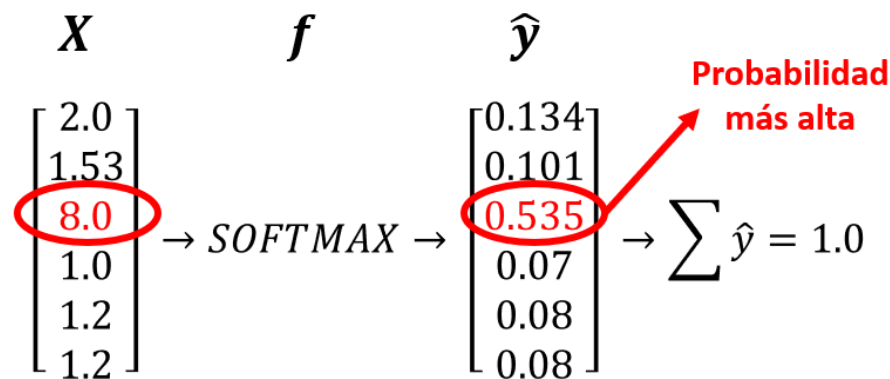


Figura 2.27: Función Softmax.

- Back Propagation

Calcula las derivadas de 'z', 'w', 'b' y 'a' para la gradiente de descenso y asigna un valor a 'w' y 'b'.

Redes Neuronales Convolucionales

Son unas redes neuronales especializadas en imágenes (Mishra, 2020). Su funcionamiento a nivel conceptual es la detección de patrones simples en las primeras capas, y a medida que se va profundizando en más capas, los patrones a detectar se vuelven más complejos.



Fuente: Curso de Especialización en Deep Learning
Andrew NG

Figura 2.28: Ejemplo de capas en las redes neuronales convolucionales.

A nivel técnico, una imagen está dividida en una matriz tridimensional, las filas y columnas conforman respectivamente el alto y ancho de la propia imagen, mientras que la profundidad de la imagen se divide en 3 capas: rojo, verde y azul (RGB).

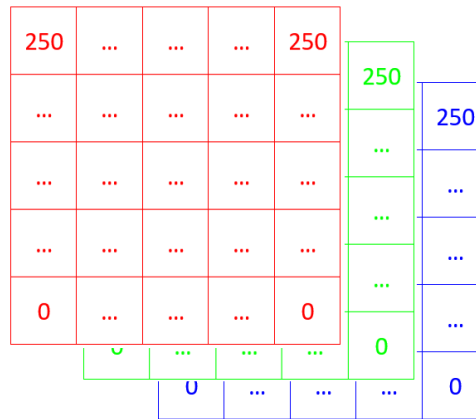


Figura 2.29: Ejemplo de una imagen a nivel técnico.

A estas matrices se les aplican transformaciones a través de filtros con distintos fines, como modificar dicha imagen o reconocer patrones y por ende objetos. A estas transformaciones se les denomina convoluciones, y no es más que la multiplicación de una submatriz de la imagen por un filtro.

Hay distintos tipos de convoluciones:

- **Valid:** No se aplica padding (relleno), por lo que no se reduce la imagen.
- **Same:** Se aplica padding (relleno), por lo que la imagen de entrada es la misma que la imagen de salida.

Tras la convolución, se debe multiplicar otra submatriz por el filtro. En este caso, por comodidad, elegimos la siguiente columna, y en caso de no existir, elegimos la siguiente fila hasta que terminemos de recorrer toda la imagen con un solo filtro. En el caso de existir más filtros, se repetiría el proceso desde la matriz original. En pro de una explicación didáctica, hemos elegido una convolución tipo *valid*, por lo que este proceso nos deja con una matriz reducida, y tantas capas como filtros se hayan aplicado.

Finalmente, existe un concepto que hemos pasado por alto, el concepto de stride (salto). Tras la multiplicación no es obligatorio escoger la siguiente columna, podemos aplicarlo en una columna sí y en otra no, al igual que en las filas, es decir, $\text{stride} = 2$; los saltos serán tan grandes como el stride que especifiquemos. Esto reducirá nuestra matriz más rápidamente, y por ende, nuestros cálculos, pero también se perderá más información.

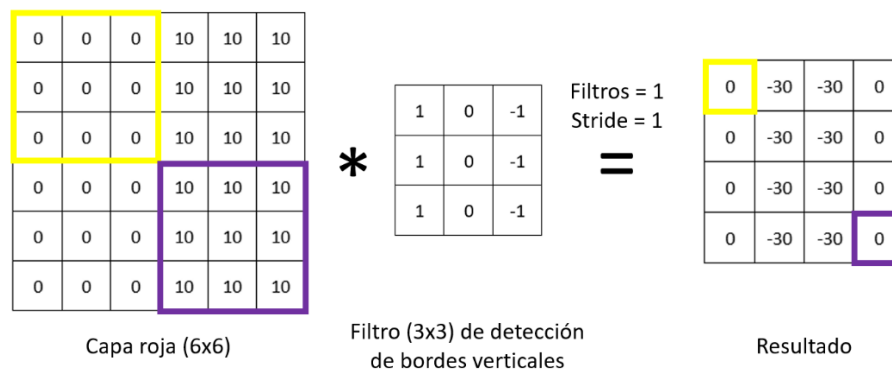


Figura 2.30: Ejemplo del procedimiento de las convoluciones básicas.

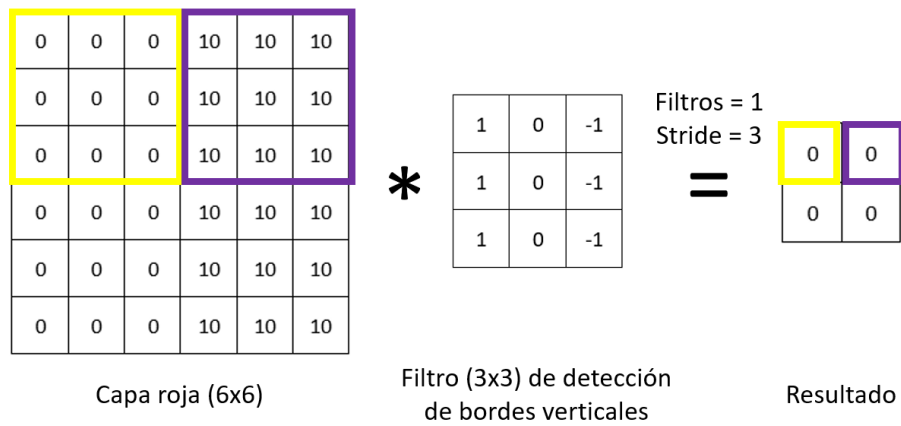


Figura 2.31: Ejemplo del procedimiento de las convoluciones con stride = 3.

Por ello, las redes convolucionales tienen distintos tipos de capas:

- **Convoluciones [CONV]:** Aplicar transformaciones en la imagen.
- **Pooling [POOL]:** Nos permite reducir el tamaño espacial de la matriz, por lo que se reduce el número de cálculos y parámetros. Los dos tipos más conocidos son:
 - Max-Pool: Elige el número máximo entre todos los números del filtro. Es el tipo de pooling más usado.
 - Avg-Pool: Hace una media entre todos los números dentro del filtro.

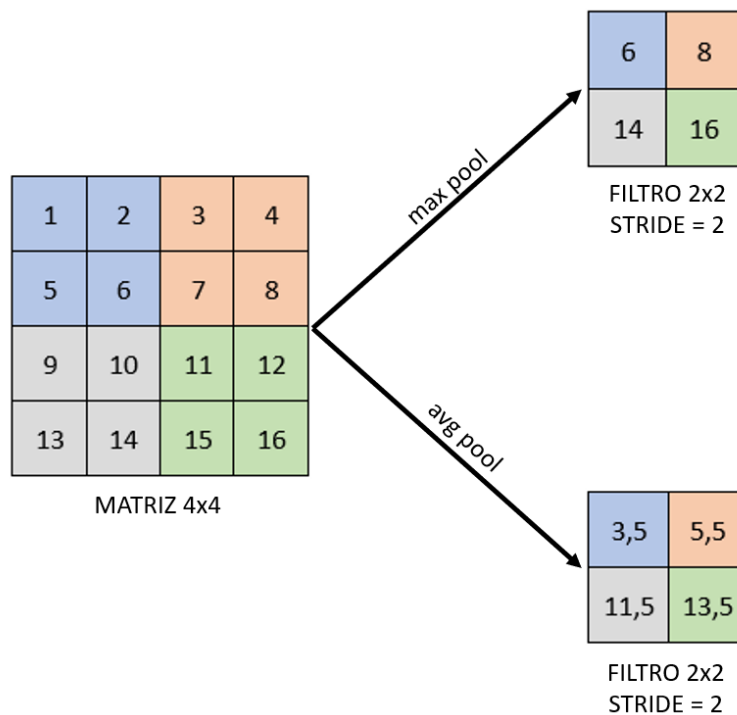


Figura 2.32: Ejemplo de distintas implementaciones de Pooling.

- **Flatten:** Pasar de una matriz bidimensional o tridimensional a un único vector.
- **Fully Connected [FC]:** Conectar todas las neuronas de la capa anterior con todas las neuronas de la siguiente capa.
- **Locally Connected [LC]:** Conectar partes locales en lugar de todas las neuronas para que en lugar de recorrer toda la matriz para encontrar un patrón, recorra una parte.

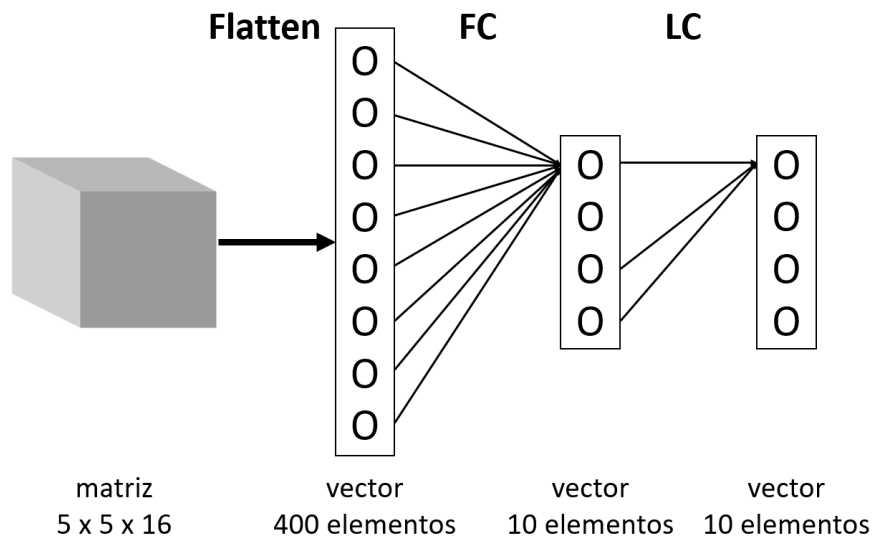


Figura 2.33: Ejemplo de vectorización de una imagen.

Redes Neuronales Recurrentes

Es un tipo especial de redes neuronales, permite guardar información a corto y largo plazo para que las posteriores iteraciones sobre la red, denominadas epochs (épocas), puedan retroalimentarse de las anteriores epochs (Phi, 2018).

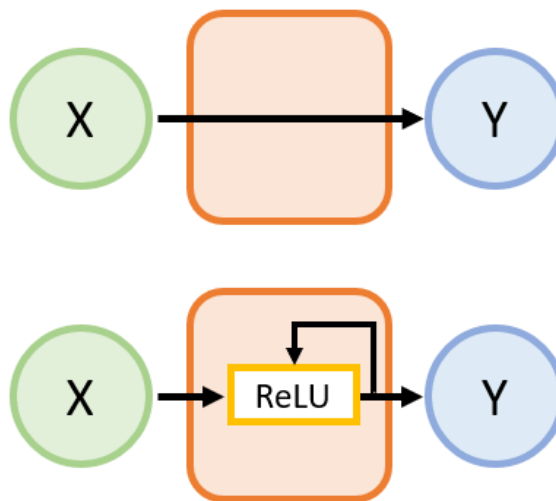


Figura 2.34: Ejemplo de red estándar (arriba) y red recurrente a corto plazo (abajo).

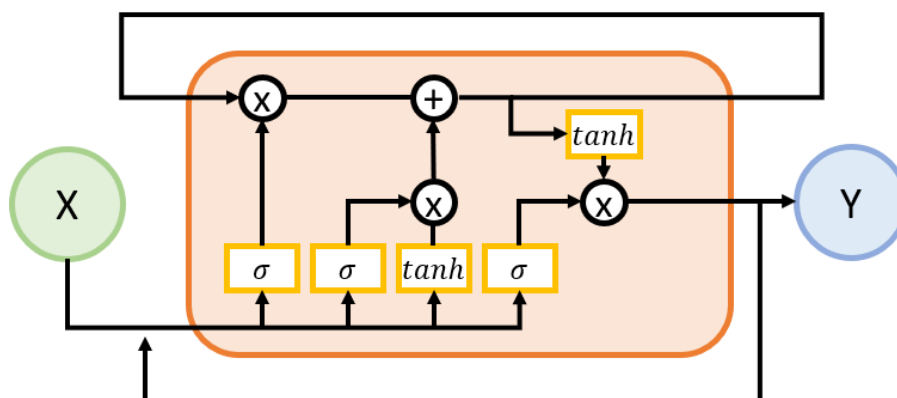


Figura 2.35: Red Neuronal Recurrente a Largo Plazo (LSTM).

2.5.2. Aplicaciones en el ámbito educativo

Aleks

Inteligencia artificial que se ha usado con 25 millones de estudiantes a la hora de enseñar matemáticas, química, estadística y contabilidad (ALEKS, 2020). Este software se especializa en determinar, de forma precisa, el conocimiento del alumno en una materia específica. Así se presentan los temarios necesarios para que siga formándose de una forma más efectiva. Esto facilita el estudio al alumno, además de personalizar cada enseñanza al estilo único que tiene cada alumno de aprender.

WriteToLearn

No hay discusión en que saber escribir y redactar es una habilidad crucial y necesaria, sobretudo a la hora de hablar del sector educativo (Pearson, 2016). Esto es lo que afirman los desarrolladores de WriteToLearn, una aplicación presentada a los profesores que permite crear una nueva tarea en la cual los alumnos deberán escribir una redacción. Esta redacción será entregada a una IA que la puntuará automáticamente fijándose en el vocabulario, la redacción, ortografía, etc.

Esto genera automáticamente informes al docente indicando dónde debería focalizar su esfuerzo, para prestar especial atención a las necesidades de la clase o a un alumno en concreto. Además, reducirá el tiempo que el docente invierte en corregir redacciones, por lo que dispondrá de más tiempo para preparar clases, dar una enseñanza más tranquila y de mejor calidad, y poder centrarse a cada alumno de manera individual.

Unicheck

Aplicación usada en el entorno universitario. Permite a los docentes comprobar de forma automática cada vez que un alumno finaliza una tarea, si contiene algún tipo de plagio o copia. Esto promueve y anima a los alumnos a que se esfuercen en crear sus propias obras.

Este tipo de software es muy conocido y Unicheck (2017) es solo un ejemplo de todos los que hay actualmente en el mercado, como por ejemplo Turnitin, Plag.es, Viper, etc.

Learning Navigator

Herramienta gratuita diseñada por Gooru (2021) que se autodefine como un GPS educativo de forma que te va guiando entre los cursos o contenido (de la propia comunidad o externos), de forma que filtra e indica el camino a seguir para llegar a la especialización deseada. De esta forma, recibes un trato personalizado que te libra de navegar entre miles de cursos y grados online.

Mangahigh

Portal enfocado en las matemáticas que permite a los alumnos de distintos niveles hasta secundaria, recibir una tarea que al ser corregida cambia el rumbo en el que se darán las futuras clases dependiendo del resultado (Mangahigh, 2008). Esto guía automáticamente a los alumnos gracias al análisis de los datos recogidos y tomando una decisión en función del nivel de la clase.

TeSLA

Es un sistema que permite verificar si la otra persona al otro lado de la pantalla durante una clase online es la que dice ser (European Commission, 2016). Para ello utiliza herramientas de reconocimiento facial y de voz, además de instrumentos que capturan patrones de escritura, detectan el plagio y analizan el lenguaje usado. Este sistema tiene como objetivo garantizar una formación en línea de mayor calidad e incrementar la reputación y credibilidad de las titulaciones online.

Thinkster Math

Es una herramienta que se encarga de analizar los pasos que sigue un alumno a la hora de resolver un problema de matemáticas, de manera que crea informes detallados sobre sus habilidades en las diferentes áreas evaluadas (Thinkster, 2015). Esto permite asignar tutores personalizados a los estudiantes de manera que puedan crear tareas individualizadas, para así trabajar sus fortalezas y debilidades.

2.5.3. Aplicaciones en el ámbito educativo

Herramientas de detección de las emociones

- iMotions

Software que mediante una serie de sensores modulares, los cuales son seleccionables, permite ver las emociones que está sintiendo una persona expuesta a un contenido como podría ser un vídeo, un anuncio, una página web, etc. (IMotions, 2006)

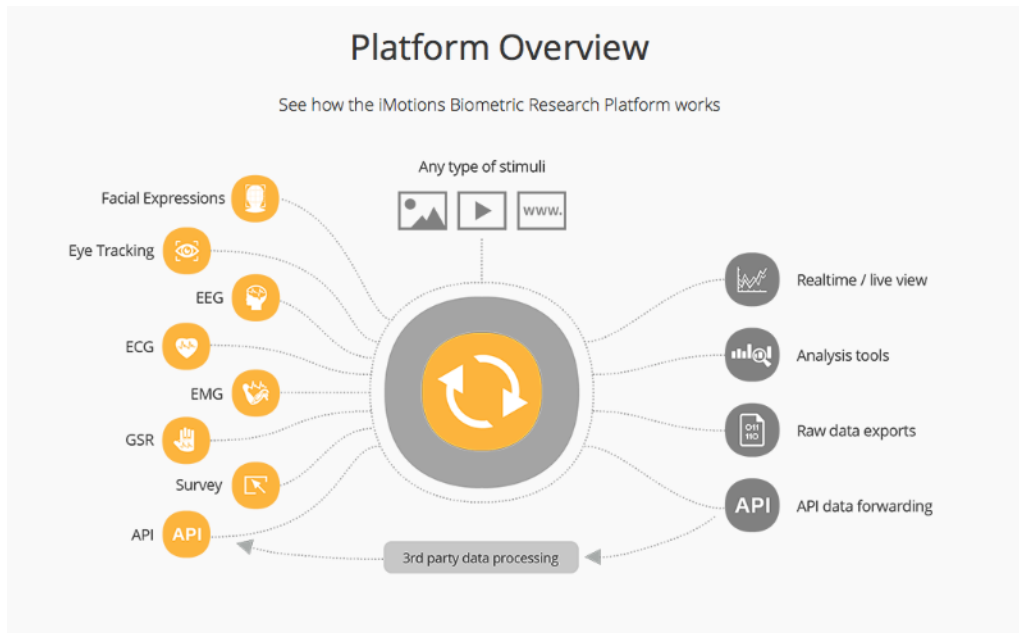


Figura 2.36: Funcionalidades de iMotions.
Fuente: (iMotions, 2006)

Este software muestra información referente a seis emociones básicas en la escala que la IA cree detectar: Tristeza, asco, felicidad, sorpresa, miedo, desprecio.

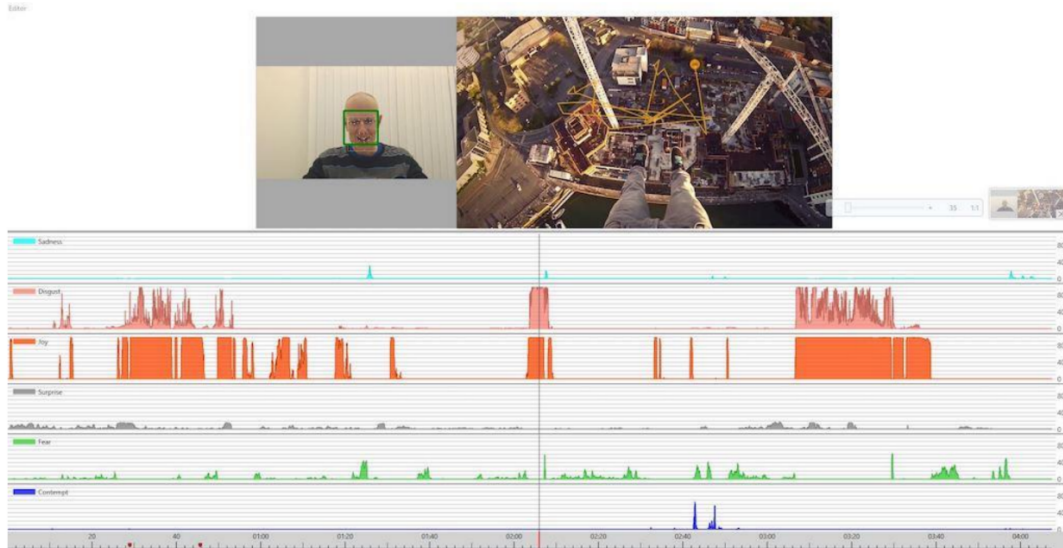


Figura 2.37: Ejemplo de uso de iMotions.
Fuente: (iMotions, 2006)

- EmotionCues

EmotionCues (Zeng et al., 2020) es un sistema analítico que se centra en analizar clases presenciales desde la perspectiva de las emociones. Al finalizar la clase, proporciona un resumen con todas las emociones captadas, de manera que el docente pueda retroceder

y hacer un seguimiento de las emociones de cada uno de sus alumnos. De esta manera el docente podrá mejorar su manera de impartir clases.

No consigue buenos resultados con oclusión de rostro, además tiene una tasa de error considerable

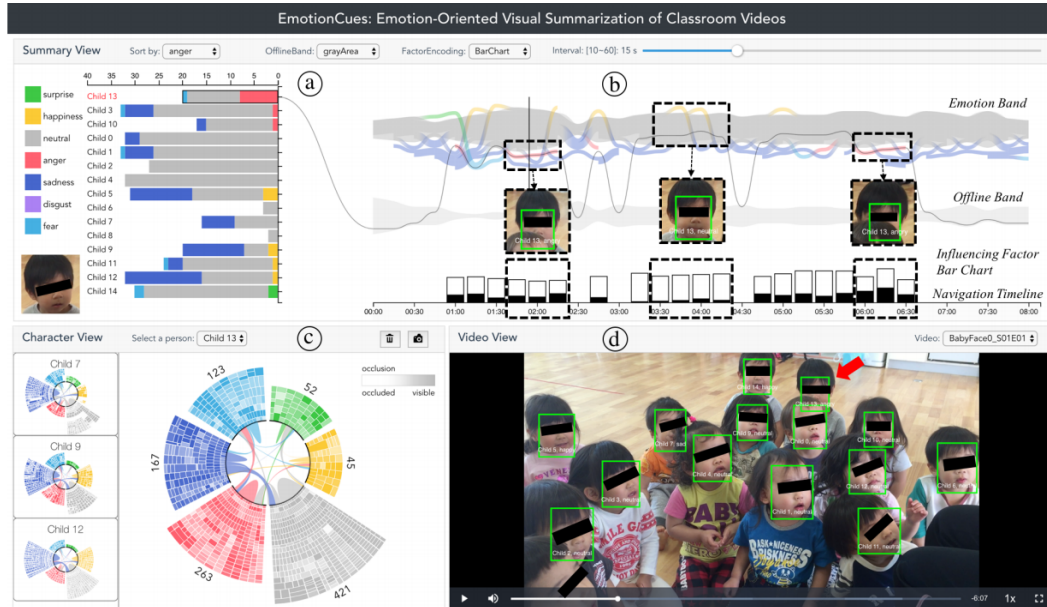


Figura 2.38: Ejemplo de uso de EmotionCues.

Fuente: (Zeng et al., 2020)

- IA de Microsoft para el reconocimiento de emociones

Microsoft (2010) ha desarrollado una herramienta para la detección de emociones llamada Face API. Esta herramienta permite detectar de manera muy precisa los sentimientos de ira, desprecio, asco, miedo, felicidad, tristeza y sorpresa.

Utiliza un algoritmo que se encarga de encontrar la cara de las personas, encuadrarlas y evaluar la emoción que se muestra en dicha imagen.

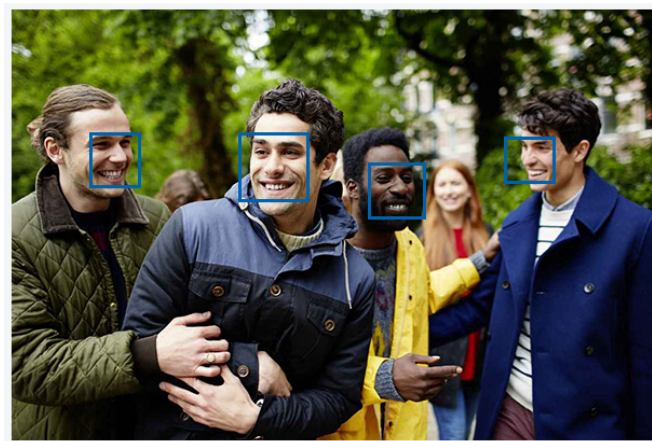


Figura 2.39: Búsqueda de rostros con el uso de Face API.

Fuente: (Microsoft, 2010)

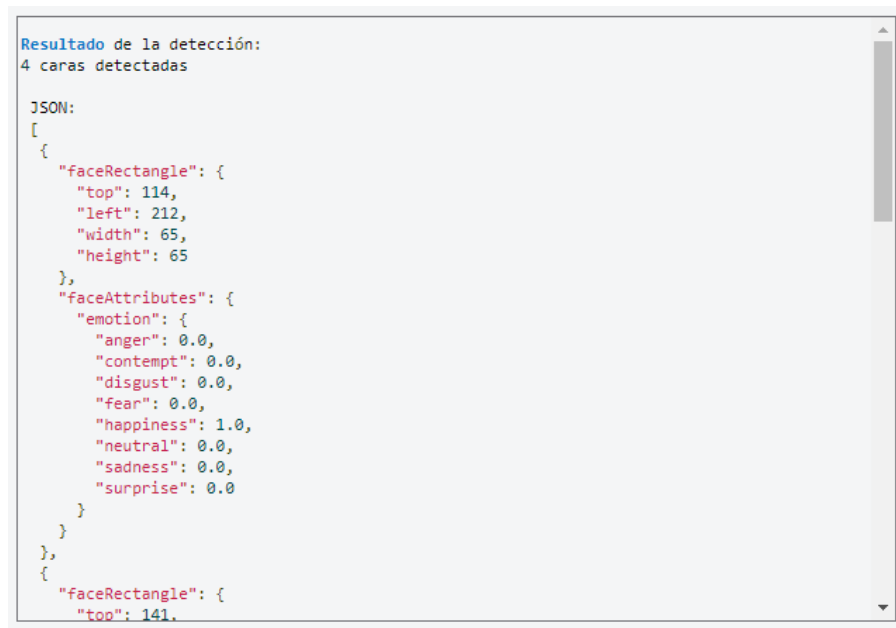


Figura 2.40: Predicciones de los rostros analizados por Face API.
Fuente: (Microsoft, 2010)

Face API también puede usarse en vídeos, de manera que reconoce las emociones de las personas que aparecen en este, y devuelve un reporte a modo de resumen con todas las emociones captadas.

- Emotient

Empresa dedicada al reconocimiento de emociones a tiempo real (Emotient, 2012). La herramienta, que esta empresa proporciona, permite identificar las emociones de alegría, sorpresa, temor, ira, desagrado y desdén.

Se puede emplear para analizar el comportamiento del usuario al ver un anuncio, a la hora de comprar o al recibir atención personalizada. De esta manera las empresas podrán mejorar sus productos y su trato con el cliente.

Herramientas de detección del aburrimiento

- Sensor del MIT y mPath para la detección del aburrimiento

Esta herramienta consiste en unos sensores colocados en los dedos índice y corazón y un aparato atado a la muñeca el cual procesa los cambios en la conductividad de la piel. La detección de picos en la conductividad se traduce como estrés y nerviosismo, mientras que las caídas de conductividad supondrán aburrimiento o desinterés.

Este sensor fue construido con la intención de ser utilizado para probar nuevos productos y experiencias, de manera que las empresas pudieran mejorar conociendo el punto exacto en el que el sujeto sintió aburrimiento (Matheson, 2017).

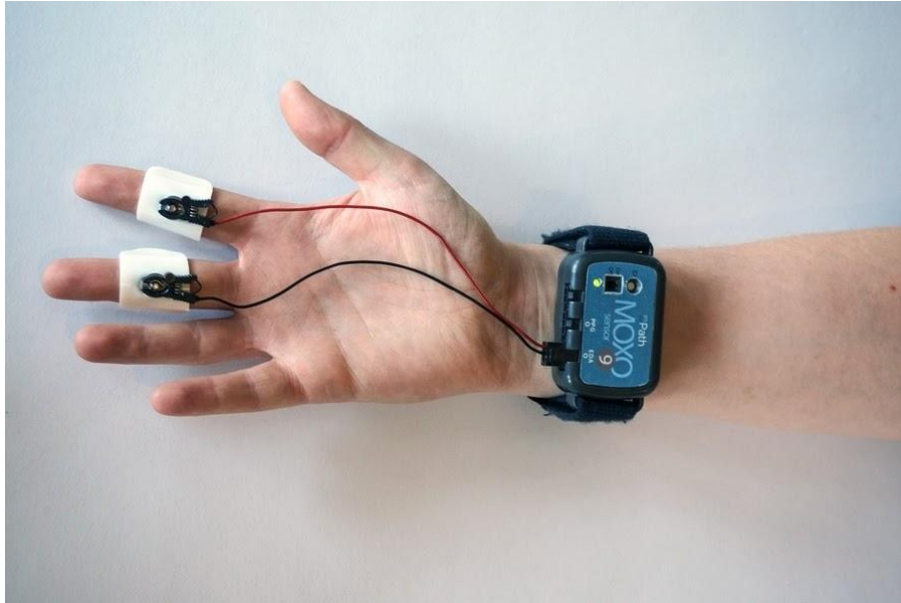


Figura 2.41: Sensor diseñado por el MIT.
Fuente: (Matheson, 2017)

- Detector de aburrimiento

Investigadores del MIT han logrado desarrollar un sensor que ayuda a detectar si la persona con la que estás hablando siente que la conversación es interesante o no (BBC, 2006). Para ello este detector está formado por unas gafas con una cámara integrada y un pequeño aparato, el cual es el encargado de procesar las imágenes y detectar las emociones. En el caso de no tener la atención de la otra persona, el software mandará una señal para que el aparato vibre en la mano del usuario.

Está centrado en un ámbito psicológico y medicinal, dado que su uso está pensado para personas con dificultades para comunicarse.

- Nestor

Sistema desarrollado por LCA para una escuela de negocios en París (Toor, 2017). Mediante el uso de una webcam, analiza el rostro y la postura de los estudiantes detectando si están atendiendo. Han afirmado que no servirá como penalización, si no que se usará para realizar preguntas a los que el sistema detecte como despistados, de manera que vuelvan a centrarse y así poder seguir el curso de la clase.

Capítulo 3

Objetivos

3.1. Objetivos y plan de trabajo

El objetivo principal de este proyecto, es el diseño de una herramienta capaz de detectar el aburrimiento en el alumnado durante una clase, además de transmitirle esta información al docente. Esta solución es demasiado compleja y extensa, por lo que en este proyecto trataremos de crear una herramienta en fase prototipo de manera que nos permita tener una primera toma de contacto con la detección del aburrimiento y su transmisión al docente. Nos centraremos en la detección facial, creando de esta manera un proyecto ampliable, el cual podría complementarse con detección postural y otros aspectos visuales que ayuden a la herramienta a cumplir su objetivo.

El proyecto se ha dividido en cuatro procesos:

1. Brainstorming / investigación de emociones: En esta etapa, Borja y Meriem nos dejaron pivotar el proyecto, para ello debíamos investigar y comprobar la viabilidad del mismo, tras investigar sobre las emociones y su detección, encontramos algunos datasets interesantes, pero hallamos pocos documentos relacionados con una emoción en específico y que a nuestro parecer era muy interesante, el aburrimiento, todas las investigaciones respecto a esta emoción tenían algo en común, eran muy invasivas de cara al usuario, ya que se debían conectar aparatos específicos.

Nuestros objetivos se centraban en :

- Investigar las emociones.
- Investigar la detección de emociones.
- Conocer el contexto de las investigaciones sobre emociones.
- Investigar campos de las emociones sin explotar.
- Investigar herramientas existentes relacionadas con las emociones.

2. Investigación del aburrimiento: En esta fase buscamos más en profundidad todo lo relacionado con el aburrimiento. Para el apartado teórico y posibles dudas, Borja y Meriem nos presentaron a Josefa Ros Velasco, experta en la materia y profesora en la Universidad Complutense de Madrid.

Nuestros objetivos en esa etapa variaron:

- Investigar el aburrimiento.
- Investigar la detección de aburrimiento.
- Conocer el contexto de las investigaciones sobre aburrimiento.
- Investigar en qué campos podríamos aplicar la detección de aburrimiento.
- Investigar herramientas existentes relacionadas con el aburrimiento.

3. Implementación de la herramienta: Tras investigar y decidir que el Trabajo de Fin de Grado lo enfocaríamos a la detección del aburrimiento, nos pusimos a crear un modelo para la arquitectura de la inteligencia artificial, y posteriormente a codificar la herramienta que nos permitiría crear el dataset, la cual hemos usado en el entrenamiento de la inteligencia artificial. Esto no supuso un gran problema ya que todos teníamos unos conocimientos básicos de inteligencia artificial.

Una vez tuvimos una base clara sobre el aburrimiento aparecieron otro tipo de objetivos que pretendíamos alcanzar con nuestra herramienta:

- La herramienta debe ser capaz de detectar, al menos, una cara a tiempo real.
- La herramienta debe ser lo menos intrusiva posible.
- La herramienta debe tener al menos un 70 % de precisión.
- Utilizar vídeos con una licencia Creative Commons (Commons, 2001), que nos permita descargarlos y modificarlos, para la puesta en marcha de nuestro experimento.
- Pasar el experimento a al menos 3 personas por integrante

4. Creación de documentos: Este proceso es el único que se solapaba con los demás, ya que la creación de la memoria ha sido durante todo el proyecto, hemos documentado toda la investigación en el apartado 2, las propuestas en el apartado 3, etc. . .

Tras conseguir los resultados en la fase de implementación, decidimos crear el paper, ya que no encontramos ninguna investigación educativa que detectase el aburrimiento a través de técnicas de aprendizaje automático.

Los objetivos en esta etapa fueron:

- Redactar el artículo **Vital Emo: The boredom detector with a machine learning perspective**, que recoja el proceso de construcción de la herramienta Vital Emo y de esta manera poder presentarlo al congreso internacional ICCE (2021).
- Redactar una memoria que recoja de manera clara todo nuestro proyecto.

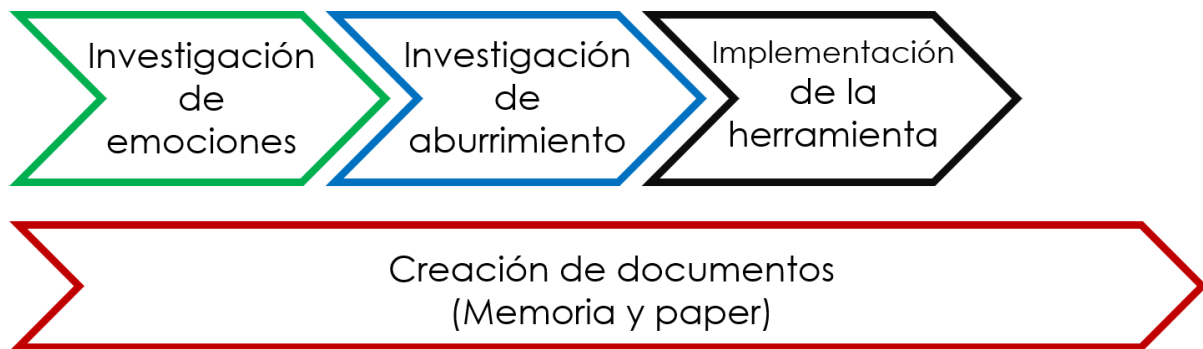


Figura 3.1: Planificación del proyecto.



Figura 3.2: Objetivos del proyecto.

Capítulo 4

Metodología

En este apartado expondremos los dos experimentos planteados. Primeramente, encontraremos el experimento que sirvió para crear el dataset, en el cual mostramos los vídeos a los participantes, grabándoles, y haciendo que puntúen el vídeo respecto a cómo se han sentido. Posteriormente, observaremos el planteamiento del segundo experimento, el cual no se llevó a cabo, haciendo uso de Amazon MTurk, en el cual los participantes califican los vídeos expuestos según su reacción emocional hacia ellos, de manera que obtuviéramos un conjunto de vídeos asociados a una etiqueta emocional.

4.1. Reacción directa al vídeo

Objetivo

El objetivo de este experimento es poder crear el dataset que se usará para entrenar a la inteligencia artificial. Este consiste en una encuesta y las grabaciones de distintos individuos, para poder asociar las caras con su estado de aburrimiento. Clasificaremos estos resultados con una escala del 1 al 9 mediante una encuesta, y todo esto se realizará de forma telemática.

Participantes

El grupo de participantes está formado por 12 personas (50 % hombres y 50 % mujeres) entre 18 y 56 años y de diferentes nacionalidades. De estas 12 personas, 5 eran estudiantes mientras que el resto tenían diferentes empleos. La titulación máxima que habían cursado los participantes iba desde la educación obligatoria hasta la universidad. La mayoría de los participantes tenían como lengua principal español, y en inglés un conocimiento suficiente como para entender una conversación o un monólogo.

Tecnología / Instrumentos

El set de vídeos contiene 19 vídeos de entre 48 segundos y 3 minutos (Badiou, 2012; Dalby, 2013; Amigo, 2014; Fortuna, 2014; Lons, 2021; Meirieu, 2019; Paniagua, 2017; Pena, 2018, 2019; Portuondo, 2019; Sotelo, 2016; Tomás Abraham, 2015), extraídos de Youtube con licencia Creative Commons (2001). Todos ellos corresponden a fragmentos de conferencias o clases salvo dos, los cuales hemos utilizado para obtener datos relacionados con notas

bajas. Los vídeos se ordenaron de forma aleatoria para cada uno de los participantes. Los vídeos se ordenaron de forma aleatoria para cada uno de los participantes.

Hemos empleado diferentes herramientas online para la grabación del experimento. Entre estas plataformas encontramos Google Meet, Discord y Microsoft Teams, seleccionando la más adecuada dependiendo de la disponibilidad y los recursos de cada participante. Por lo que los experimentos se realizaron de manera telemática.

A cada participante se le avisaba de que se iba a comenzar la grabación y se le explicaba en qué consistía el experimento, además de leerle un acuerdo legal de consentimiento de participación en el experimento. Una vez hecho esto, se rellenaba un cuestionario con los datos personales como el nombre, nivel de estudios, idiomas hablados, etc.

Posteriormente, cada participante visualizaba un vídeo y respondía unas preguntas sobre el mismo. Este último proceso se repetía para cada uno hasta consumirlos todos, o hasta que la persona quisiese parar o descansar.

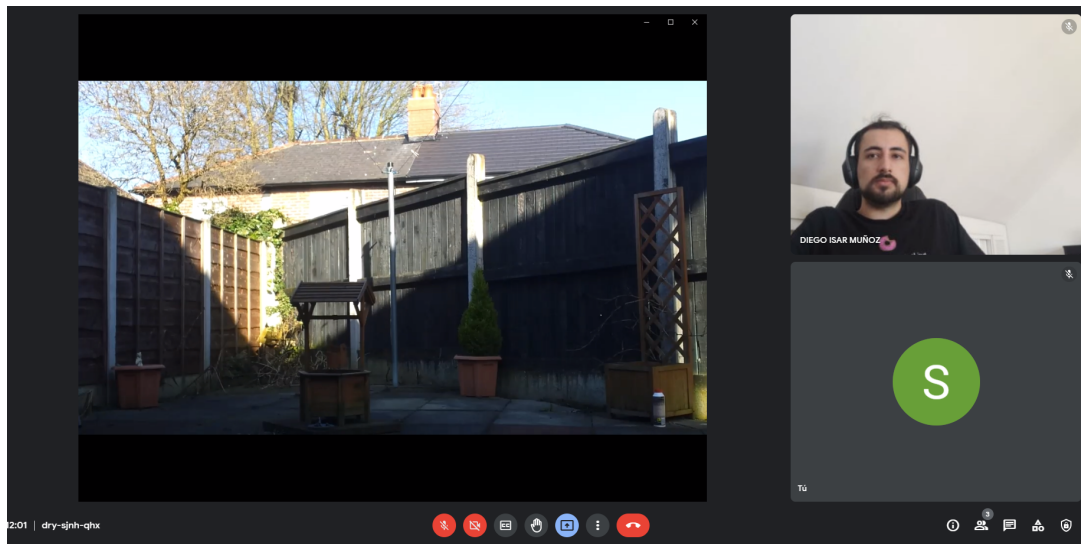


Figura 4.1: Ejemplo de experimento simulado.

Toma de resultados

Los dos cuestionarios fueron creados con la herramienta Google Forms. El primer cuestionario está destinado a la recogida de datos personales pre-experimento, de manera que podamos obtener estadísticas basándonos en la información recopilada. El segundo cuestionario es post-experimento y está destinado a la recogida de datos del estado de aburrimiento tras ver cada uno de los videos mostrados.

Las encuestas eran las siguientes:

Variable a medir	Tipo	Rango
Nombre	Texto	-
Género	Opción	Mujer, Hombre, Otro
Edad	Número	18 - 99
Nacionalidad	Texto	-
Ocupación	Texto	-
Idiomas	Texto	-
Estudios	Opción	Primaria, ESO, Bachillerato, FP, Universidad, Máster y Doctorado

Cuadro 4.1: Encuesta de datos personales.

Variable a medir	Tipo	Rango
Aburrimiento	Escala	1 (aburrido) - 9 (interesante)
Seguir viendo	Si o No	0 - 1

Cuadro 4.2: Encuesta de calificación de vídeos.

4.2. Amazon MTurk

Es una herramienta que permite, a través de una página web, realizar tareas que requieren de inteligencia humana (HIT). A estas personas se les denomina Turkers, y trabajan a cambio de una compensación económica. Amazon MTurk permite:

- Procesar fotos y vídeos
- Verificar y limpiar datos
- Recopilar información
- Procesar datos

Requester	Title	HITS	Reward	Created	Actions
Amazon Requester Inc. - C	[French language proficiency requir...	61,046	\$0.01	17h ago	Preview Accept & Work
Amazon Requester Inc. - C	[日本語能力が必要]商品のインタレ...	59,647	\$0.01	7h ago	Preview Accept & Work
Amazon Requester Inc. - C	Product to Interest Audit (single yes/...	28,379	\$0.01	1h ago	Preview Accept & Work
Amazon Requester Inc. - C	[dominio del idioma español requeri...	27,670	\$0.01	21h ago	Preview Accept & Work
Amazon Requester Inc. - C	[Proficiência no idioma português br...	19,719	\$0.01	20h ago	Preview Accept & Work
Crowdsurf Support	Transcribe up to 35 Seconds of Med...	17,485	\$0.05	3m ago	Preview Quality
TC Research	Find the Email for These Mental He...	13,896	\$0.12	5d ago	Preview Accept & Work
UnSpun Opinions	Opinion Survey	12,180	\$0.50	1m ago	Preview Accept & Work

Figura 4.2: Trabajos ofrecidos para Turkers en Amazon MTurk.
Fuente: Amazon (2005)

Objetivo

El objetivo es conseguir un dataset que nos permita clasificar un vídeo en las categorías de interesante (I), neutral (N) o aburrido (A), y asignar a dicha clasificación un valor numérico comprendido entre 0 y 1, siendo el sumatorio de todas las categorías igual a 1.

Vídeo	Aburrido (A)	Neutral (N)	Interesante (I)
Vídeo 1	0.1	0.1	0.8
Vídeo 2	0.5	0.3	0.2

Cuadro 4.3: Ejemplo de clasificación.

Tecnología

Para poner en marcha nuestro experimento en Amazon MTurk, primero debemos recortar partes de conferencias que, a nuestro juicio, sean aburridas o interesantes. A lo largo de todo el trabajo, dichas conferencias deben tener una licencia de Creative Commons que nos permita modificar, en este caso recortar, partes de estas. Tras ello, subiremos dichos recortes a la plataforma y los trabajadores de Amazon MTurk deberán etiquetar los vídeos. Para ello, hemos añadido un cuestionario que nos permite clasificar dichos vídeos en las categorías de ‘interesante’, ‘neutral’ o ‘aburrido’. El cuestionario es el siguiente:

PUNTÚA DEL 1 AL 9 QUE TE HA PARECIDO EL VÍDEO, SIENDO EL 1 MUY ABURRIDO, Y SIENDO EL 9 MUY INTERESANTE

ABURRIDO

INTERESANTE

1	2	3	4	5	6	7	8	9

Cuadro 4.4: Casillas de la encuesta en Amazon MTurk.

Los pesos que hemos asignado a cada celda son los siguientes:

99 % A	75 % A	50 % A	25 % A	0 % A	0 % A	0 % A	0 % A	0 % A
0 % N	0 % N	25 % N	50 % N	99 % N	50 % N	25 % N	0 % N	0 % N
0 % I	0 % I	0 % I	0 % I	0 % I	25 % I	75 % I	75 % I	99 % I

Cuadro 4.5: Distribución de pesos Amazon MTurk.

El uso de estos pesos es estrictamente por cuestiones éticas, ya que nuestra Inteligencia Artificial nunca va a poder estar al 100 % segura de que una persona esté aburrida o interesada. Esta distribución permite que la clasificación tenga el mismo peso, esto es necesario dado que el algoritmo de Inteligencia Artificial, si tuviera un peso distinto, en este caso en el estado neutro, clasificaría como neutro la mayoría de estados, pudiendo arrojar falsos positivos y falsos negativos en los otros dos estados.

Para obtener nuestro dataset, debemos juntar todas las respuestas y calcular los pesos con:

$$\frac{\sum_{i=1}^n \text{probabilidad}}{\sum_{e=1}^3 \sum_{i=1}^n \text{probabilidad}_{ei}}$$

Por ejemplo:

- Usuario 1 vota la casilla 2 para el vídeo 1
 - Aburrido: 75 %
 - Neutro: 0 %
 - Interesante: 0 %
- Usuario 2 vota la casilla 3 para el vídeo 1
 - Aburrido: 50 %
 - Neutro: 25 %
 - Interesante: 0 %
- Usuario 3 vota la casilla 7 para el vídeo 1
 - Aburrido: 0 %
 - Neutro: 25 %
 - Interesante: 50 %

$$\text{Aburrido} = \frac{75 + 50 + 0}{(75 + 50 + 0) + (0 + 25 + 25) + (0 + 0 + 50)} = 0,56$$

$$\text{Neutro} = \frac{0 + 25 + 25}{(75 + 50 + 0) + (0 + 25 + 25) + (0 + 0 + 50)} = 0,22$$

$$\text{Interesante} = \frac{0 + 0 + 50}{(75 + 50 + 0) + (0 + 25 + 25) + (0 + 0 + 50)} = 0,22$$

Vídeo	Aburrido (A)	Neutral (N)	Interesante(I)
Vídeo 1	0.56	0.22	0.22

Cuadro 4.6: Diseño experimental de Amazon MTurk.

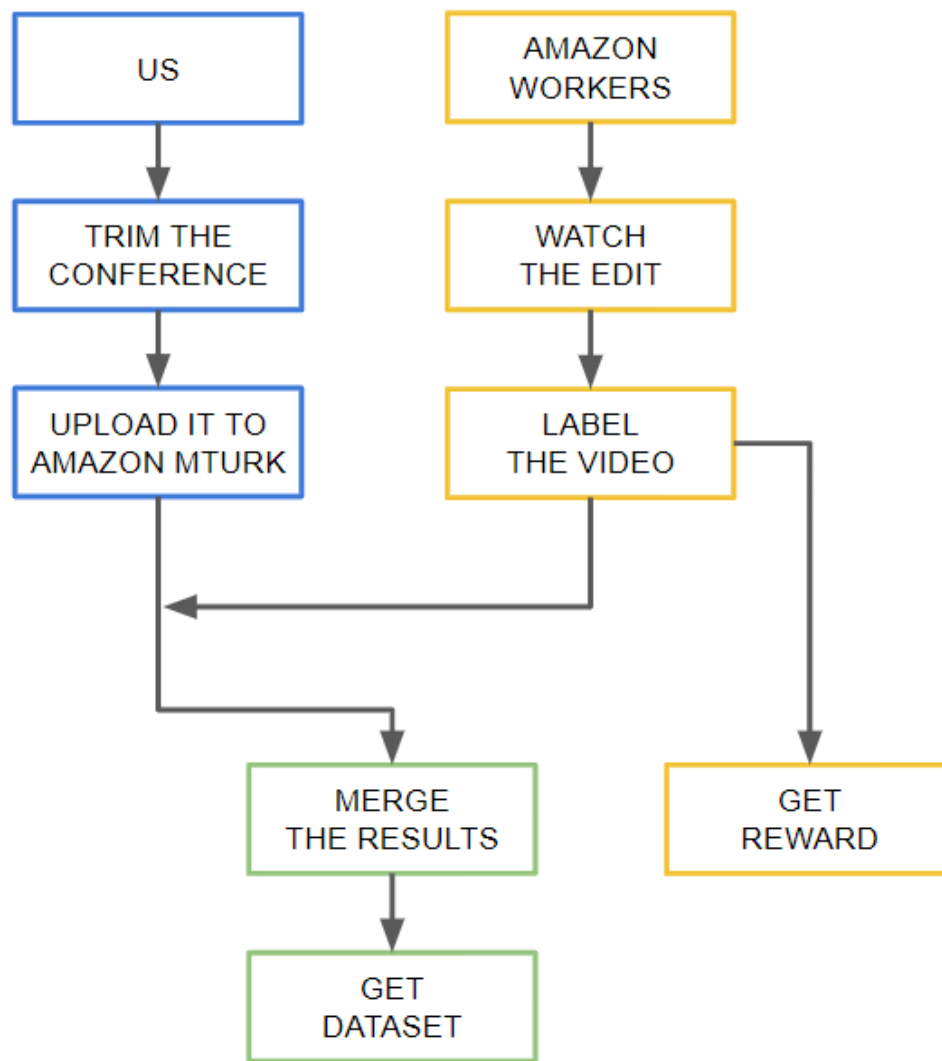


Figura 4.3: Diseño experimental de Amazon MTurk.

Al plantear el anterior experimento, nos dimos cuenta de que había 2 razones principales por las que no se pudo llevar a cabo:

- Nuestros vídeos son en castellano, por lo cual, no nos interesaba que una persona que no supiese el idioma puntuase los vídeos. Esto nos lleva a establecer filtros de idioma, y Amazon, por cada filtro, obligaba a pagar una comisión de aproximadamente \$1 más por cada participante, por lo que conseguir un dataset con una alta fiabilidad superaba el presupuesto disponible.
- No a todas las personas les parece interesante el mismo tipo de vídeos, si puntuamos los vídeos y no las caras corremos el riesgo de que nuestra Inteligencia Artificial detecte patrones que no están asociados al aburrimiento o al interés.

Por estas razones, se desechó el experimento con MTurk y solo se realizó finalmente el experimento de personas reaccionando directamente al vídeo.

Capítulo 5

Arquitectura

Para abordar el problema descrito desde una perspectiva tecnológica y siendo lo menos intrusivos posible con el usuario, decidimos hacer una detección del aburrimiento a través de reconocimiento facial. Esto es, a través de una cámara que capta la expresión facial del usuario y, usando inteligencia artificial, detectamos patrones relacionados con el aburrimiento.

El modelo de inteligencia artificial elegido es una red neuronal convolucional, que nos permiten clasificar una imagen por clases a través de la repetición de un patrón. Por ejemplo, una red neuronal convolucional es capaz de detectar una cara a través de patrones como los ojos o la boca.

En este caso buscamos que sea capaz de detectar expresiones faciales relacionadas con el aburrimiento o engagement. Concretamente, hemos planteado un modelo basado en VGG16 donde las imágenes de entrada a la red neuronal son de un tamaño de 48X48 píxeles en escala de grises. A estas imágenes se les aplica el siguiente patrón de capas tres veces:

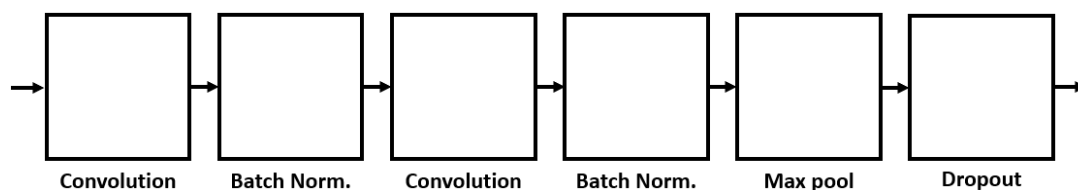


Figura 5.1: Patrón de capas utilizado

Después de aplicar este patrón, la información pasa por dos capas Fully Connected. La capa de salida genera una salida de 7 clases: enfado, asco, miedo, felicidad, tristeza, sorpresa y neutral.

Tras el entrenamiento, aplicamos Transfer Learning (Sarkar, 2018). Para ello hemos re-seteado los valores de las 10 últimas capas y, en la última, hemos establecido tres clases (aburrimiento, neutral e interesado) en lugar de las siete clases de salida anteriores. Finalmente reentrenamos las 10 últimas capas de la red neuronal.

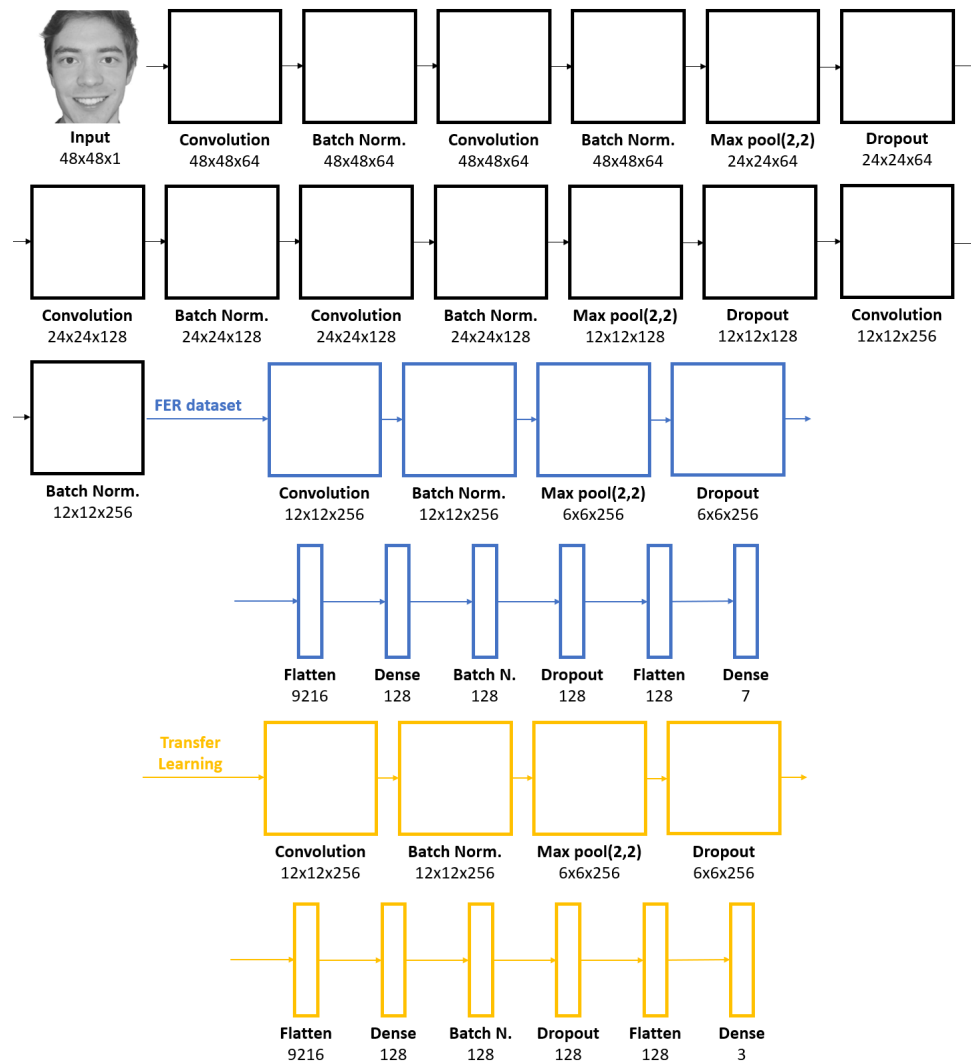


Figura 5.2: Arquitectura del proyecto.

Las redes neuronales están compuestas por nodos que simulan el comportamiento del cerebro humano. Estos nodos están conectados entre sí mediante valores denominados pesos, y se denomina entrenamiento a la variación de dichos pesos con el objetivo de reducir el error de convertir una entrada a una salida. Un caso específico son las redes neuronales convolucionales, que a través de matrices más pequeñas y distintas denominadas filtros, encuentran patrones en una imagen.

Un caso específico son las redes neuronales convolucionales, que a través de matrices más pequeñas y distintas denominadas filtros, encuentran patrones en una imagen.

Las imágenes que hemos usado para entrenar la inteligencia artificial han sido extraídas de dos datasets: dataset de Kaggle y un dataset propio.

Facial Expression Recognition(FER) Challenge de Kaggle es un dataset que nos ha servido de apoyo para reconocer caras y emociones a partir de imágenes y sus respectivas etiquetas. Este dataset tiene aproximadamente 36.000 filas, y está dividido en las columnas:

- *emotion*: Número entero que representa la emoción que se expresa en la imagen, siendo 0 = Angry, 1 = Disgust, 2 = Fear, 3 = Happy, 4 = Sad, 5 = Surprise, y 6 = Neutral.
- *usage*: Training, Public Test y Private Test.
- *pixels*: 2304 números enteros del 0 al 255. Representan un píxel en escala de grises.

Hemos ampliado este dataset para tener más datos de entrenamiento usando técnicas de Data Augmentation (i.e. rotación, desplazamiento, zoom) mediante la API de Keras.

Para el entrenamiento, el dataset se ha dividido en un 90 % train set y un 10 % test set. Estos datos nos sirven como base para que la red convolucional sea capaz de detectar caras y pequeños rasgos de emociones.

En la investigación que hemos realizado, no hemos encontrado un dataset público que diferencie específicamente entre aburrimiento, estado neutral e interés (engagement). Por ese motivo, hemos decidido crear nuestro propio dataset. En el apartado ‘Metodología’ explicamos más en detalle el procedimiento que hemos seguido para obtener los vídeos del dataset. Cada uno de estos vídeos ha tenido un procesamiento posterior:

- a). División del vídeo en frames.
- b). Detección y recorte de caras.
- c). Conversión a escala de grises.
- d). Reducción del tamaño a 48x48x1 utilizando OpenCV.
- e). Generación de archivo CSV con el identificador del usuario, el identificador del vídeo, el identificador del frame, los píxeles y la puntuación de engagement que ha sentido en el momento de ver el vídeo (proporcionada por el usuario).

Como cada miembro del equipo debía encargarse de la grabación y posterior transformación del vídeo a datos, llegamos a la conclusión de que lo mejor sería crear un programa entre los cuatro para automatizar la tarea. Por otro lado, al no ser un sistema perfecto, detectamos que OpenCV, en algunas ocasiones, en lugar de detectar una cara, detectaba partes del fondo de la habitación o camisetas, por lo que decidimos exportar las imágenes a una carpeta y revisarlas, en el caso de encontrar alguna anomalía, eliminamos la fila correspondiente.

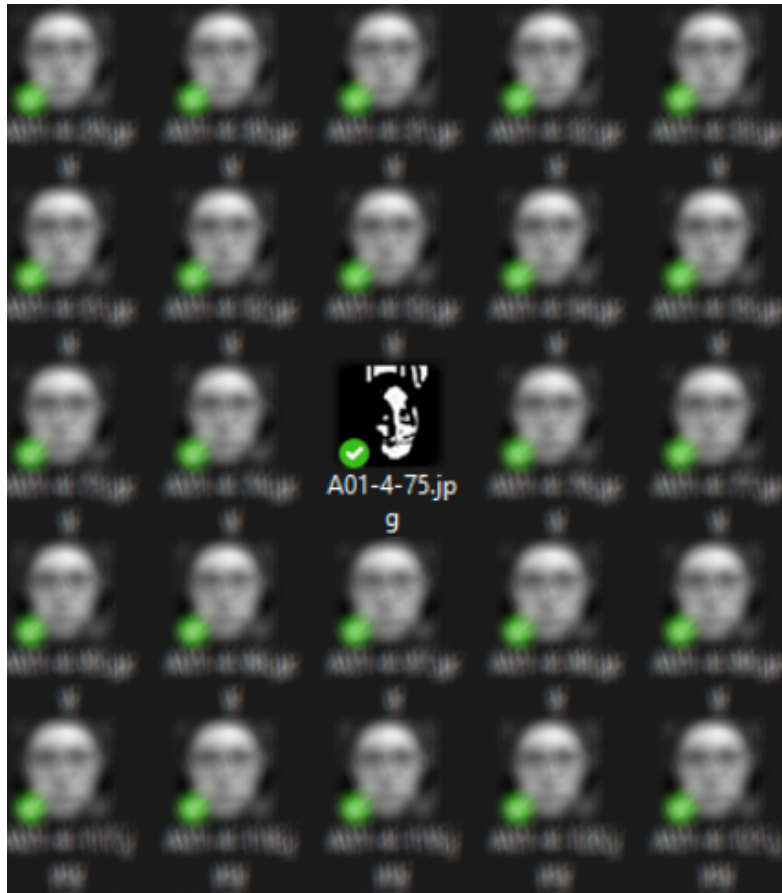


Figura 5.3: Imagen anómala detectada.

A01,4,71,76	75	73	64	58	47
A01,4,72,76	75	73	67	58	46
A01,4,73,75	75	72	65	58	45
A01,4,74,76	76	75	74	66	60
A01,4,75,75	73	73	68	61	51
A01,4,76,76	74	73	72	61	54
A01,4,77,75	73	73	66	59	47
A01,4,78,76	73	73	71	61	52
A01,4,79,76	73	74	71	60	52

Figura 5.4: Línea anómala.

El dataset se ha dividido por el identificador de la persona. Esto se ha hecho así porque si se mezclaran todos los frames, podría proporcionar falsos positivos en la fase de validación que, posteriormente, no tendrían que corresponder con los valores de acierto en un entorno real. La división de este dataset ha sido de un 70 % de personas a entrenamiento y un 30 % de personas a cross validation.

Capítulo 6

Resultados

Tras grabar a todos los participantes durante el experimento y procesar los vídeos, se han obtenido aproximadamente 64.000 frames de 12 participantes distintos junto su correspondiente puntuación.

Con el cuestionario que se realizó posteriormente, se obtuvieron 226 respuestas totales, cuyos valores han resultado distribuidos entre las diferentes opciones del cuestionario como se ve en la Figura 3. Como anticipamos, hay un mayor número de puntuaciones con valores 1 (aburrido) y valores 7 y 8 (valores muy cercanos a interesado), confirmando que la selección de vídeos que hemos realizado produce en el participante emociones más cercanas a los límites que neutras.



Figura 6.1: Resultado de las puntuaciones de los vídeos.

Los resultados obtenidos de la segunda pregunta del cuestionario muestran una distribución casi equitativa entre los vídeos valorados de forma positiva y negativa (56.2 % respondieron que sí querían continuar viendo el vídeo, frente a un 43.8 % que no).

Emociones

Una vez finalizado el entrenamiento del clasificador, con el conjunto de datos que etiqueta las emociones obtenido de Kaggle, se puede observar que según avanzan las iteraciones de

entrenamiento, va mejorando la precisión llegando a mantenerse en un 70 %; mientras que el coste va disminuyendo, como se puede observar en la Figura 6.2 y la Figura 6.3.

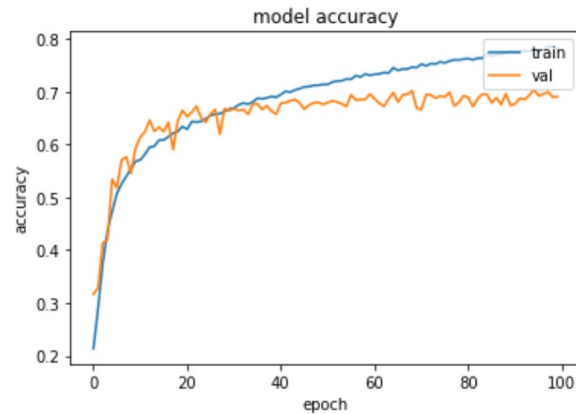


Figura 6.2: Precisión del modelo de las emociones.

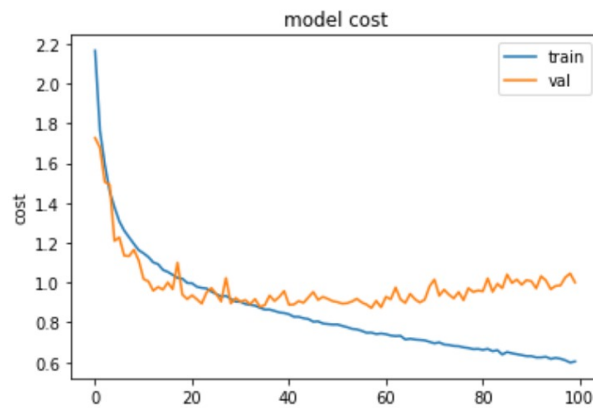


Figura 6.3: Coste del modelo de las emociones.

Esto nos permite que las primeras capas del clasificador obtengan suficiente información para detectar rasgos faciales como la boca, nariz, ojos o incluso gestos faciales como una sonrisa o una mueca.

Tras realizar pruebas, se puede apreciar que el clasificador funciona en tiempo real de forma muy precisa. Los rostros que expresan alegría, neutral, sorpresa, miedo o tristeza son los que detectan con mayor facilidad mientras que el enfado y el asco son las emociones que más fallan a la hora de ser clasificadas.

Aburrimiento

Una vez obtenido estos resultados, se eliminaron las capas correspondientes y se añadieron las nuevas. Se utilizó el dataset recogido por nosotros para reentrenar únicamente las últimas capas añadidas y, como podemos ver en las figuras Figura 6.4 y Figura 6.5, la precisión de los datos de entrenamiento aumenta mientras que los datos de validación permanecen estáticos entre el 30 % y 40 % de precisión.

Vemos que lo que ha ocurrido es un efecto de sobreajuste de la red neuronal, ya que pese a que la entrada consta de muchos frames etiquetados, gran parte de ellos son muy

similares entre sí (al ser del mismo individuo).

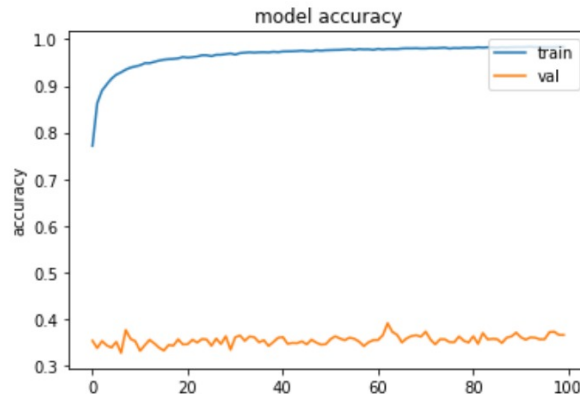


Figura 6.4: Precisión del modelo del aburrimiento.

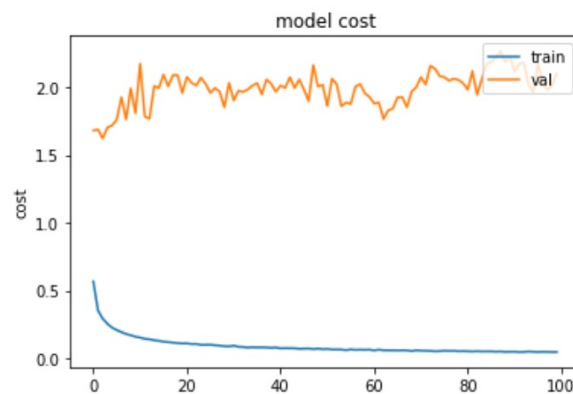


Figura 6.5: Coste del modelo del aburrimiento.

Sin embargo, el clasificador sí consigue predecir ciertos patrones concretos que son síntomas del aburrimiento o de interés. Durante el proceso de pruebas, se ha podido comprobar que la aplicación reacciona indicando aburrimiento, cuando se produce una oclusión de la cara causada por la mano, el sujeto entrecierra los ojos o gira la cabeza respecto a cámara que está grabando. En contraparte, podemos apreciar cómo otros gestos, como esbozar una sonrisa, tener los ojos completamente abiertos o estar directamente en frente a la cámara es definido como interés.

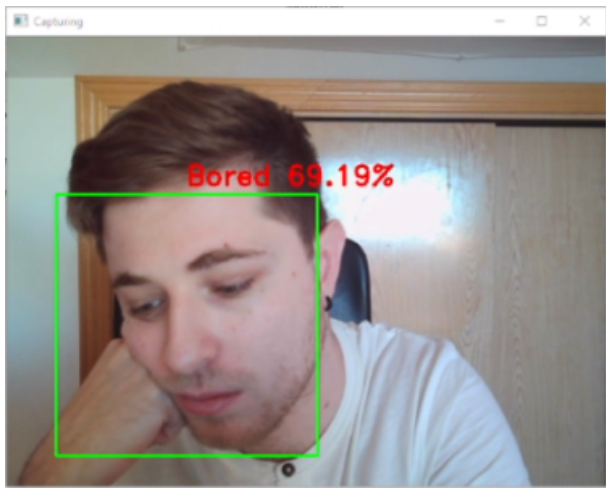


Figura 6.6: Detección facial y su porcentaje de aburrimiento.

También se ha adaptado la herramienta para su uso en plataformas de videollamada online, en este caso se ha probado en una reunión de Google Meet, donde capturamos la pantalla, analizando las caras que contiene, y mostrando en la parte superior izquierda el conteo de estados que se detectan.

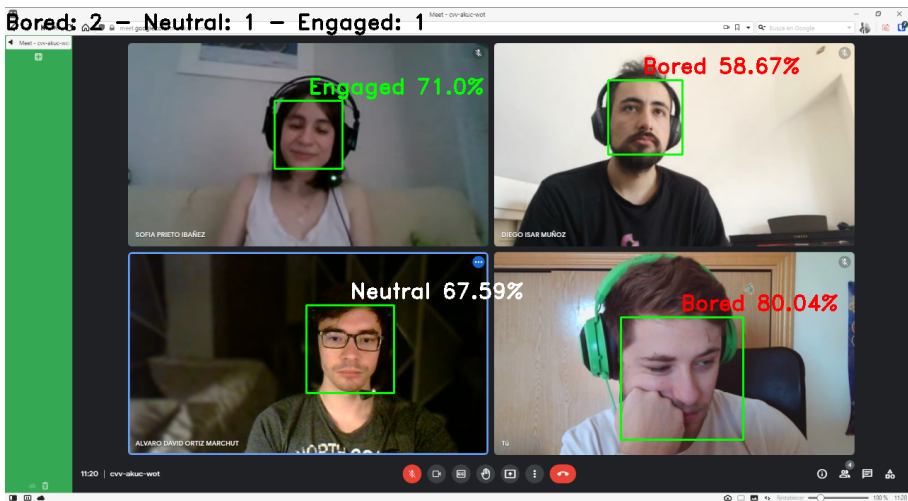


Figura 6.7: Ejemplo de uso real en una reunión.

Tras múltiples entrenamientos, se observa cómo los patrones detectados son variables, dependiendo de los datasets seleccionados para el entrenamiento, y la validación de los resultados. A pesar de esto, algunos patrones como sujetarse la cabeza o esbozar una sonrisa, son consistentes tras varios entrenamientos.

La idea principal del entrenamiento era reutilizar el concepto de pesos explicado en la sección de Amazon MTurk:

99 % A	75 % A	50 % A	25 % A	0 % A	0 % A	0 % A	0 % A	0 % A
0 % N	0 % N	25 % N	50 % N	99 % N	50 % N	25 % N	0 % N	0 % N
0 % I	0 % I	0 % I	0 % I	0 % I	25 % I	75 % I	75 % I	99 % I

Cuadro 6.1: Distribución de pesos Amazon MTurk.

El problema de esta distribución es que, al ser tan concreta y al tener un dataset con pocas caras, la inteligencia artificial no predecía correctamente casi ningún resultado.

La diferencia, por ejemplo, de la clase 2 a la clase 3 es relativamente pequeña y al tener 9 clases, se contabilizaba como error, por lo que la Inteligencia Artificial no aprendía correctamente ningún patrón facial. La solución fue unir las 9 clases en 3 clases.

Aburrimiento	Neutralidad	Interés
---------------------	--------------------	----------------

Cuadro 6.2: distribución usada.

Los porcentajes de acierto que se muestran son la predicción de la función softmax de la última capa.

Al probar la herramienta en un caso real, donde se deben detectar múltiples caras, se puede observar que a mayor número de rostros, aumenta el coste computacional ralentizando la tasa de actualización de la imagen.

Capítulo 7

Limitaciones del estudio

Como se puede observar en los resultados, la aplicación no ha quedado lista para su uso en un entorno real. Esto se debe a ciertas limitaciones que se han podido apreciar y que se han sufrido a la hora de trabajar:

- **Los únicos datos que se analizan y que se obtienen del sujeto es su rostro.** Esto limita el aprendizaje, ya que el lenguaje corporal y/o un detector más avanzado podrían haber dado más información relevante.
- **La información analizada es estática.** Los datos son únicamente imágenes individuales, de modo que no se puede detectar patrones o comportamientos que indicarían un estado de aburrimiento, como podrían ser un rápido pestañeo, un bostezo largo, una pérdida del contacto visual prolongada. Al tener exclusivamente imágenes aisladas, lo único que se puede analizar serían datos atemporales, como que la persona tiene los ojos cerrados o su boca está abierta.
- **Poca variedad en el dataset.** Durante el experimento se ha conseguido grabar el rostro de 12 personas viendo 19 vídeos cada uno. Esto ha proporcionado muchas imágenes de rostros, pero con poca variedad; y ha causado un sobreajuste en la inteligencia artificial.
- **Un desconocimiento en el campo.** Debido al perfil técnico de los integrantes del grupo, a pesar de haber contactado con profesionales en el ámbito, no contábamos con un alumno que tuviera formación específica en el campo del aburrimiento. Esto supuso que gran parte del tiempo inicial del proyecto se dedicase a la investigación y a la parte teórica, por lo que tuvimos menor tiempo para la parte práctica del experimento.
- **Material hardware insuficiente.** Dado que en el proyecto se maneja un gran volumen de datos, y una arquitectura de inteligencia artificial de cómputo costoso, la falta de potencia de cálculo ralentizó tanto la iteración sobre los datos, como el entrenamiento sobre el modelo.
- **Poca variedad de vídeos.** Ha sido escasa la diversidad de conferencias que se disponían, debido a las restricciones de derechos de autor para su uso en el experimento. Era necesario el uso de conferencias y clases grabadas, las cuales se pudiesen utilizar para fines no comerciales, descargar y modificar.

Capítulo 8

Conclusiones

Este trabajo ha tenido como objetivo la investigación del aburrimiento y la creación de una IA, para facilitar la educación online a los docentes, dándoles feedback sobre si los estudiantes se aburren o no. Al hablar de aspectos técnicos de la construcción de la IA, en primera instancia, se entrenó haciendo uso del dataset de Kaggle Challenge (2018), el cual asocia caras recortadas con emociones. Una vez tuvimos la base que reconocía patrones asociados a las emociones, modificamos las últimas capas para así obtener la salida esperada (Aburrido, Neutro, Interesado), y se volvieron a entrenar con los nuevos datos del dataset creado *ad hoc* para este trabajo.

En este segundo entrenamiento, se utilizaron como entrada las caras obtenidas del experimento, en el cual, se mostraban vídeos cortos al participante y tras su visualización, éste asignaba una puntuación en una escala de aburrimiento.

Los puntos destacables que hemos podido observar son:

1. Nuestra inteligencia artificial clasifica a partir de gestos explícitos, i.e los rostros de las personas que se exponen al vídeo son suficientes para detectar patrones característicos del aburrimiento o interés.

Hemos podido observar como nuestra IA detecta parcialmente algunos de los patrones característicos del aburrimiento, anteriormente mencionados en el estado del arte:

- Oclusión de la cara con la mano. La herramienta creada detecta con mayor eficacia la oclusión de rostro con la mano al utilizar la mano izquierda. Esto puede ser debido a que, al tener un mayor número de personas que utilizan la mano izquierda, frente a gente que utiliza la mano derecha.
- El sujeto entrecierra los ojos.
- Girar la cabeza respecto a la cámara que está grabando.

También podemos observar patrones de interés, los cuales detecta parcialmente nuestra IA:

- Esbozar una sonrisa.
- Tener los ojos completamente abiertos.

- Estar directamente en frente a la cámara.

Pero existen algunos patrones que nuestra IA no logra detectar o no está preparada para su detección. Entre ellos encontramos:

- El ensanchamiento de los párpados. Actualmente logra detectar la apertura de estos, pero no su ensanchamiento.
- La apertura de los labios. Solo detecta la apertura para expresar sonrisas, no identifica los bostezos u otra clase de apertura.
- Posturas, como mantenerse erguido. La herramienta no está preparada para la detección de patrones corporales, ya que se centra simplemente en gestos faciales.

2. Hemos tenido en cuenta que este proyecto, donde utilizamos inteligencia artificial, requiere una cantidad de datos mucho mayor que la obtenida. Aun así, con los rostros de únicamente 12 personas, hemos conseguido unos resultados esperanzadores respecto a llevar este proyecto a una mayor escala. Hemos conseguido un 40 % de índice de acierto sobre el propio dataset, pudiendo mejorar al introducir más variedad entre los rostros recogidos, ya que actualmente, aunque tengamos muchas imágenes, son todas muy parecidas, y nuestra IA se sobreajusta con el dataset que entrena, produciéndose peores resultados al validarla. Además de conseguir más personas, se debe tratar siempre de conseguir diversidad entre los datos, como contar con diferentes nacionalidades, edades, culturas, etc; para evitar en mayor medida un posible sesgo en la IA.

8.1. Futuras líneas de trabajo/actuación

Con este trabajo realizado, se ha conseguido el esqueleto de lo que sería el proyecto completo, pero queda trabajo para conseguir una aplicación totalmente funcional. Los puntos principales que habría que desarrollar son:

- **Analizar más datos visuales que complementan al rostro.** El cuerpo humano expresa el aburrimiento de diferentes formas. Utiliza además de la cara, elementos como la postura, la dilatación de las pupilas, etc. Sería necesario incorporar nuevos módulos, los cuales permitan la recogida y transformación de esos datos, así como su entrenamiento.
- **Implementación de un análisis en el tiempo.** Utilizando redes recurrentes, en vez de un análisis estático de una única imagen, se podría conseguir que la inteligencia artificial pueda detectar patrones, como podría ser un tic nervioso, una distracción constante o una disminución de la frecuencia del parpadeo. Esto ayudaría, en gran medida, a clasificar el estado de una persona con mayor seguridad.
- **Aumentar la diversidad y el tamaño de los datos.** Un proyecto de esta complejidad requiere un tamaño de datos mayor. Es necesaria una manera más eficaz y rápida de recoger datos, como por ejemplo, grabar a todos los alumnos de una clase o realizar y facilitar a los individuos una aplicación, la cual sea capaz de grabar automáticamente al reproducir el vídeo.

■ **Disponer de una variedad de competencias en el equipo de desarrollo.**

Las emociones, concretamente el aburrimiento, es un campo complejo y desconocido para los integrantes del equipo. Aunque se ha recibido ayuda de profesionales el campo, sería necesario disponer, de forma activa, de una persona instruida en el ámbito para facilitar el desarrollo y la investigación.

Son muchas las futuras opciones y el trabajo que quedaría por delante para finalizar el proyecto. Durante el desarrollo se ha creado un prototipo de herramienta que podría complementarse, de modo que, teniendo el análisis facial, se podría ir ampliando de forma incremental con otras extensiones que afiancen el análisis y la detección del aburrimiento.

Chapter 8

Conclusions

The main objective of this project is to investigate the state of boredom and create an AI, to facilitate online education for teachers, giving them feedback on how bored are the students during an online class. If we focus on the technical aspects of the AI construction, in the first place, it was trained using the Kaggle Challenge (2018) dataset, which associates cut-out faces with emotions. Once we had the base that recognized patterns associated with emotions, we modified the last layers in order to obtain the expected output (Bored, Neutral, Interested). To finalize, the last layers were retrained with the new data.

In the second training, we used as input the faces obtained from the experiment, in which the person was shown short videos and after viewing them, assigned a score on a boredom scale.

The highlights we have been able to observe are:

1. Our artificial intelligence classifies from explicit gestures, i.e. the faces of people exposed to the video are sufficient to detect patterns characteristic of boredom or interest.

We have been able to observe how our AI partially detects some of the characteristic patterns of boredom, previously mentioned in the state of the art:

- Face occlusion with the hand. The tool created detects face occlusion with the hand more effectively when using the left hand. This may be because we have a greater number of people who use the left hand, against people who use the right hand.
- The subject squints.
- Turn the head away from the camera that is recording.

We can also observe patterns of interest, which our AI partially detects:

- To smile.
- Eyes wide open.
- Being directly in front of the camera.

But there are some patterns that our AI fails to detect or is not prepared to detect. Among them we find:

- The widening of the eyelids. Currently, it is able to detect the opening of the eyelids, but not their widening.
 - The opening of the lips. It only detects smiling, but it does not identify yawning or any other kind of opening.
 - Postures, such as standing upright. The tool is not prepared for the detection of body patterns, as it focuses simply on facial gestures.
2. It has been taken into account that this project, where we use artificial intelligence, requires a much larger amount of data than the obtained. However, with only the faces of 12 people, we have achieved promising results that show that this project can be taken to a larger scale. We have achieved a 40% accuracy rate on the dataset itself, and we could improve it by introducing more variety among the faces collected, since currently, although we have many images, they are all very similar, and our AI is over-fitting with the dataset it trains, producing worse results when validating it. In addition to getting more people, we should always try to achieve diversity among the data, such as having different nationalities, ages, cultures, etc., to avoid possible bias in the AI.

Capítulo 9

Contribución

9.1. Diego Isar Muñoz

Vital emo nació de la vaga idea de mezclar el campo de la inteligencia artificial con la subjetividad de las emociones humanas. Principalmente, era conseguir detectar las emociones y sacar alguna aplicación sobre eso. Por lo que la primera etapa consistió en una lluvia de ideas y, contando con todo el equipo, buscar en muchos ámbitos para así poder concretar este trabajo. Mi parte consistió en la investigación del lenguaje corporal, su importancia en la detección de emociones y su posible detección automática. La investigación no fue muy fructífera, ya que no existen muchos estudios corroborados sobre el lenguaje corporal, y menos aún de la detección automática. Lo más interesante que encontré fue un guante con sensores que, mediante aprendizaje automático, era capaz de reconocer gestos.

Posteriormente, y después de reunir todas las investigaciones, el proyecto derivó a detectar el aburrimiento en una persona. Este cambio de paradigma hizo que tuviéramos que investigar también sobre el aburrimiento en una mayor profundidad, en contraste con las emociones. A consecuencia de esto, Borja y Meriem, nuestros tutores, nos pusieron en contacto con Josefa Ros Velasco, una experta en el sector del aburrimiento. Ella nos dio mucho material para investigar y poder llevar a cabo el proyecto. Mi investigación consistió principalmente en un estudio para detectar, mediante escáner encefalográfico, si había correlación entre distintas regiones del cerebro con los distintos estados de interés o desconexión, en concreto en la ínsula. En dicho experimento, pudieron encontrar que sí existía correlación, habiendo más actividad en las tareas más entretenidas. A pesar de los buenos resultados, a nosotros no nos servía, ya que uno de los objetivos propuestos es que fuera la herramienta lo menos intrusiva posible, y no existe tecnología actualmente a nuestro alcance que sea un MRI que no sea tan invasivo.

También se me encargó diseñar y plantear el experimento, que consistió en crear el dataset para poder entrenar la inteligencia artificial de nuestro proyecto, además de ofrecer ayuda en el otro experimento a Álvaro Ortiz. Para crear este experimento, tuve que atenerme a dos cosas principalmente: a que fuera lo menos intrusivo posible, tanto el propio experimento como quienes lo supervisan para no influenciar en él; y que pudiera ser telemático, debido a las restricciones del Covid-19. En primera instancia, planteé hacer una herramienta para el experimento, pero resultó más sencillo y eficiente simplemente grabar directamente las personas, y preguntar posteriormente las preguntas a cada vídeo correspondiente.

Ya una vez juntadas todo el material investigado y ya diseñado el experimento, empezamos a trabajar sobre la inteligencia artificial. Comencé a diseñar lo que sería la primera inteligencia artificial para el Transfer Learning, con el dataset de Kaggle. A pesar de que en el temario del grado en ingeniería del software no se imparte apenas conocimientos sobre inteligencia artificial, nuestro equipo estuvo suficientemente familiarizado para poder llevar a cabo esta herramienta.

En paralelo a la creación de la herramienta, estuvimos buscando los vídeos que se usarían en el experimento de creación del dataset. Fue una tarea compartida por todos los integrantes del equipo, para conseguir variedad en los vídeos. Yo me centré en buscar vídeos que no fueran muy entretenidos, desde que el temario no fuera muy entretenido o que este no se expusiera adecuadamente, o problemas técnicos como mal audio o con estática, vídeos de mala calidad, etc. A continuación elegimos los vídeos, en concreto

qué fragmentos de vídeos íbamos a emplear, y nos pusimos a cortarlos Álvaro Ortiz y yo.

Una vez obtenidos los vídeos, empezamos a iniciar el experimento para la recogida de datos, usada posteriormente para el entrenamiento de nuestra herramienta. Contribuí personalmente con 3 personas, habiendo visto cada uno un total de 19 vídeos. Sumándose al resto de la recolección del equipo, obtuvimos un total de 226 fragmentos de vídeos y 64000 frames.

Ya teniendo un buen número de datos, procedimos a usar el anterior modelo creado y aplicar Transfer Learning al nuevo modelo. Esta vez Daniel Padilla fue quien se encargó de hacer el modelo definitivo, mientras que el resto del equipo le apoyaba. Decidimos trabajar de esta manera, ya que esta parte era complicada sacar tareas para trabajar en paralelo, además de que Daniel Padilla era el único con equipo suficientemente potente como para entrenar la inteligencia artificial. Yo sobre todo aporte en el diseño definitivo del modelo, y algunas ideas para intentar eliminar el sobreajuste que apareció una vez la inteligencia artificial fue entrenada.

En la memoria, en el capítulo del estado del arte, cada uno se encargó de escribir su parte de investigación, por lo que yo escribí todo lo relacionado con el lenguaje corporal, además de la parte sobre el aburrimiento. En el resto de la memoria, repartimos por capítulos. En una primera instancia, yo redacté parte de la metodología, en concreto la parte experimental, parte de los resultados del proyecto y los objetivos, aunque este último fue reescrito por Álvaro Ortiz. Además, aporté en varias secciones, como en las líneas futuras del proyecto, y revisé la parte en inglés de la memoria. En la parte final de la memoria, revise toda la redacción, con el fin de buscar posibles errores.

Artículo Vital Emo: *The boredom detector with a machine learning perspective*

Durante el desarrollo de la herramienta, nuestros tutores nos propusieron la oportunidad de presentar este proyecto en una conferencia mundial, en el ICCE 2021 (International Conference on Computers in Education). Para poder presentarlo, debíamos preparar un paper, por lo que presentar nuestra memoria del TFG sin ningún cambio no nos valía (ver Anexo 2).

Para el paper, me centré en los objetivos del proyecto, que durante el desarrollo del paper, se convirtió en una parte de la introducción. También contribuí en las conclusiones del proyecto y en el abstract del proyecto. Además, junto a Sofía Prieto, nos encargamos de las referencias de todo el proyecto y a traducir todo el paper a inglés.

9.2. Álvaro David Ortiz Marchut

En la primera etapa, la investigación de emociones y brainstorm, decidí buscar datasets que nos facilitasen el trabajo en un futuro, para ello hice uso de Google Scholar y Kaggle. En Google Scholar busqué papers relacionados, que compartiesen los datasets usados o si eran públicos que los referenciasen, por contraparte, en Kaggle decidí buscar datasets por palabras clave como: “emotions”, “Galvanik Skin Response”, “Face tracking”, “Detector”, etc. . .

En esta etapa, encontré datasets que relacionaban constantes vitales y emociones. Por otro lado, también encontramos papers en los que se usaba Transfer Learning para aprender los patrones más básicos, por ejemplo, lo que es una cara.

No todos los datasets eran públicos, por lo que para estudiarlos en profundidad, necesitábamos hacer peticiones a entidades académicas, como es el caso de DEAP, propiedad de la Universidad de Cambridge.

Aunque no es un dataset en formato CSV, encontramos un dataset de imágenes denominado ‘IAPS’, en el que una imagen va asociada a una emoción. Estas imágenes no se pueden compartir por internet, ya que no deben ser conocidas para transmitir dicha emoción por primera vez, solo se puede pedir para proyectos de investigación o trabajos académicos.

Tras la primera iteración de esta fase, Daniel Padilla y yo investigamos sobre el Face Tracking, y encontramos una tesis doctoral que utilizaba redes neuronales convolucionales y recurrentes. Este concepto nos daría la idea de trabajo a futuro (No la hemos podido implementar), que solucionaba varios problemas que teníamos, y es que, al trabajar de forma ‘temporal’ podíamos basarnos en frames anteriores para diferenciar si una persona estaba parpadeando o durmiendo, o incluso la postura corporal que tenía. Además, investigamos sobre papers muy conocidos como es el caso de los modelos VGG16, DeepFace o FaceNet.

Tras ello, hicimos un pequeño prototipo que nos serviría para desarrollar la herramienta. Este prototipo consistía en detectar una cara, recortarla, pasarla a escala de grises, y mostrarla en pantalla.

En la segunda etapa, la investigación del aburrimiento, nos pusimos en contacto con Josefa, que nos facilitó papers que ella consideraba relevantes para nuestro proyecto, en este caso nos dividimos los papers, y tras ello hicimos una puesta en común, comentando los conceptos más importantes de cada uno. Cabe destacar la dificultad de estos papers por ahondar en conceptos técnicos, en un campo que no es el nuestro.

De forma paralela, Diego Isar y yo planteamos el diseño de los experimentos. Como al principio teníamos dos experimentos, el de Amazon MTurk y la recolección de caras, decidimos dividirnos el trabajo, él se encargó del experimento de recolección, y yo del MTurk.

En la tercera etapa, la implementación del proyecto, la empezamos con la detección de emociones. Daniel Padilla programaba en su ordenador, mientras los demás buscábamos soluciones a problemas. En un principio, la idea era que todos lo hiciéramos sobre distintas ramas del proyecto, pero esto fue altamente ineficiente porque perdíamos bastante tiempo en juntar las ramas. Otra idea era programar todos en Live, con una herramienta de Visual

Studio, el problema de esta solución era que el único que podía correr el programa era que compartía la sesión. Entonces, decidimos quien programaba de una manera muy simple, el que tuviese la tarjeta gráfica más potente, ya que teníamos que iterar bastantes veces para probar la precisión de la inteligencia artificial.

Tras varios días de programar y pruebas, decidimos empezar a recolectar los vídeos que íbamos a mostrar a los trabajadores de Amazon MTurk, para ello decidimos entrar en YouTube y buscar conferencias que tuviesen una licencia Creative Commons que nos permitiese descargar y modificar el vídeo (recortarlo).

Nos pusimos a investigar la herramienta de Amazon MTurk, y decidimos descartar la idea por los motivos expuestos en el apartado 4, tras plantear y aplicar otro experimento para crear nuestro propio dataset, grabamos las caras de nuestros familiares y amigos, y finalmente recortamos todos los vídeos.

Para automatizar el trabajo, decidimos reutilizar el recortador de caras que creamos Daniel y yo, de forma que iterábamos en todos los vídeos de una carpeta y recortábamos las caras. Esto nos permitió guardar en un dataset todas las imágenes en píxeles. Además, guardábamos las imágenes en otra carpeta, de esta forma detectábamos las imágenes anómalas y las eliminábamos. Finalmente, unimos todas las líneas de los dataset generados en un único archivo.

Cuando teníamos todo preparado, nos pusimos a implementar la inteligencia artificial como tal. Con el modelo propuesto en la arquitectura, hicimos el mismo procedimiento que para entrenar el dataset de emociones. Daniel Padilla se encargaba de programar, y nosotros de buscar información y soluciones a los problemas que se iban presentando, siendo el mismo procedimiento para el Transfer Learning.

El test de la herramienta, al no necesitar entrenamiento, lo hizo cada uno por su cuenta, con personas ajenas al proyecto que en todo momento eran informadas del procedimiento y de la poca precisión que habíamos obtenido. En este momento llegamos a las conclusiones principales, que sí había detectado indicios de aburrimiento o de interés, pero no era muy preciso.

Un mes antes de entregar la memoria, se nos presentó la oportunidad de presentar el paper y no la desaprovechamos. Ciertamente fue bastante estresante por la entrega de varias prácticas y la proximidad de los exámenes, pero mereció la pena. Todos los integrantes del equipo escribimos una parte de dicho paper, y Borja, Meriem y Josefa validaban lo que poníamos. Sin su ayuda es probable que el paper no hubiese salido.

Respecto al desarrollo de la memoria, cada persona se empezó encargando de un subapartado del apartado 2 ya que realmente estábamos investigando el ‘Estado del Arte’. Posteriormente escribimos todos los apartados en orden numérico.

Todos hemos escrito apartados y subapartados de la memoria y todos hemos corregido la memoria. Mi trabajo más destacado ha sido la redacción del subapartado de Inteligencia Artificial en el apartado 2 de esta memoria, además de la creación de las figuras de casi toda la memoria.

9.3. Daniel Padilla Rodríguez

Durante el comienzo del proyecto, la investigación de emociones, me centré en buscar investigaciones y proyectos existentes. En ese proceso, encontré una investigación titulada “Respiration-based emotion recognition with deep learning”, en la que se explicaba cómo mediante un sensor de respiración y dos regresiones logísticas predecían entre cuatro emociones. Esta investigación utilizaba el dataset de IADS, por lo que Álvaro Ortiz y yo nos pusimos en contacto con la entidad correspondiente, para ver si nos lo podían facilitar, lo cual conseguimos aunque no le llegamos a dar uso.

Después de esto, junto con Sofía Prieto, encontramos un producto llamado iMotions que, mediante unos sensores y reconocimiento facial, predecía las emociones que sentía un espectador al exponerle un vídeo en concreto. Intentamos obtener este producto para ver su funcionamiento y estudiarlo, pero al ponernos en contacto con la empresa que lo gestiona, la única opción que nos daban era adquirirlo a un precio de estudiante, el cual se salía de presupuesto, por lo que rechazamos la oferta.

Este proceso nos sirvió como inspiración para lo que se terminó convirtiendo el proyecto. Después de esto, empecé a investigar sobre la detección y el reconocimiento facial. Estudiamos varios algoritmos de detección facial, como Viola-Jones, y bibliotecas que lo implementan, como OpenFace u OpenCV, la cual se acabó usando en el proyecto final. Junto con Álvaro Ortiz, estuvimos haciendo pruebas codificando un programa simple que, mediante la webcam, cogiese la imagen, la transformase a blanco y negro, para que más adelante se detectasen las caras que contenía, y las recortase permitiéndonos almacenarlas o utilizarlas en un algoritmo futuro. Este prototipo, mucho más adelante, terminó convirtiéndose en lo que actualmente es el detector de aburrimiento y la herramienta que utilizamos con los vídeos para recortar y crear el dataset que sirvió de entrenamiento.

Con un código funcional que detectaba caras, comencé a buscar datasets que relacionasen caras con la atención, el aburrimiento o el interés que expresaban. Al no encontrar ninguno, ya que es un campo que realmente no se ha explorado a gran escala con reconocimiento facial, empecé a buscar datasets que relacionasen rostros con algún estado que nos pudiese ser de utilidad. En ese momento, encontré el dataset del concurso de Kaggle “Challenges in Representation Learning: Facial Expression Recognition Challenge”, que relaciona rostros de personas con las emociones enfado, asco, miedo, felicidad, tristeza, sorpresa y neutral. Junto con la ayuda de Álvaro Ortiz, descargamos las imágenes etiquetadas y estuvimos codificando un programa que nos permitiese ver las imágenes con sus respectivas etiquetas, para comprobar la calidad del dataset llegando a la conclusión de que nos sería útil.

Debido a que yo tengo el hardware más potente dentro de los integrantes del grupo, me dediqué a investigar e instalar todo lo necesario para utilizar TensorFlow usando mi GPU en vez de CPU. Esto nos facilitó mucho a la hora de entrenar e iterar en el desarrollo de la Inteligencia Artificial, debido a que aceleró en gran medida el proceso de entrenamiento. Mientras tanto, mis compañeros se encargaron de buscar información y documentación sobre qué arquitectura nos era más conveniente usar.

Durante la codificación de la Inteligencia Artificial, participamos todos los integrantes del grupo. Implementamos diferentes tipos de arquitectura que propusieron mis compañeros, y las entrenamos con el dataset de emociones obtenido en Kaggle. Una vez realizado esto,

unimos el detector facial que codifiqué anteriormente para realizar pruebas en un entorno real, detectando las emociones que expresaba nuestro rostro.

Teniendo una base, nos pusimos todos los integrantes del equipo a buscar vídeos para crear nuestro dataset de aburrimiento. Para ello, entre todos, buscamos en YouTube vídeos de conferencias o clases que tuviesen una licencia que nos permitiese su uso, descargar y modificar. Tras comparar los vídeos encontrados, seleccionamos los intervalos que íbamos a usar para mostrar a la gente mientras grabábamos su rostro para crear el dataset.

Como la idea principal del proyecto era conseguir que, mediante Amazon Mechanical Turk, nos etiquetaran los vídeos según lo interesante que les parecía, todo el grupo investigamos las tecnologías que usaba, precios e intentamos generar una tarea de ejemplo para, próximamente, usarlo en un caso real. Finalmente, debido al precio extra que suponía añadir una restricción para que la gente que nos puntuase los vídeos hablase español, ya que la plataforma funciona por defecto en inglés, se descartó esta idea y se optó por facilitar, a la gente que era grabada, un formulario.

Para la recogida del dataset, participamos todos los integrantes con la ayuda de conocidos y familiares que se prestaron a ver los vídeos y ser grabados dando su consentimiento. Yo grabé a tres de las doce personas que se presentaron voluntarias, además de ayudar a Sofía Prieto con la grabación de otras cuatro personas diferentes, debido a problemas técnicos.

Una vez recogido todo el material para crear el dataset, procedí a recortar los rostros en frames de los vídeos que yo había grabado junto a sus puntuaciones. Una vez acabé con mis vídeos, ayudé a mis compañeros con sus vídeos y a editar algunos vídeos que daban problemas a la hora de detectar las caras, debido a la iluminación y fondos confusos que producían falsos positivos.

Teniendo ya un dataset funcional y con el modelo que detectaba emociones, procedimos todos a realizar transfer learning y probar a entrenar con multitud de parámetros diferentes, para ver cual nos daba mejores resultados. Además, probamos con diferentes personas, luminosidad, webcams, ordenadores, etc. Para el apartado de la memoria de resultados, me encargué de sacar imágenes de las estadísticas de los datos obtenidos y gráficas del aprendizaje del modelo. Además, junto con Álvaro Ortiz, fui el encargado de pasar toda la memoria a Latex y darle formato.

Artículo Vital Emo: *The boredom detector with a machine learning perspective*

Decidimos redactar un paper para el 29th International Conference on Computers in Education, donde exponíamos toda nuestra investigación.

Durante el trabajo de redacción, yo me centré en analizar y redactar los apartados de resultados, limitaciones, conclusiones y futuro trabajo (ver Anexo 2) que finalmente se entregaron.

9.4. Sofía Prieto Ibáñez

El proyecto comenzó con una etapa de brainstorming, en la que debatimos cómo enfocar el tema del TFG. Para tener una visión más clara de las posibilidades a las que podíamos optar, cada integrante investigó el tema del reconocimiento de emociones desde diferentes perspectivas. En mi caso, opté por la investigación de las emociones desde el punto de vista de los EEG (Electroencefalogramas), ya que me resultaba muy interesante la combinación de la IA con la medicina. Entre las múltiples investigaciones existentes, encontré papers esperanzadores con altos porcentajes de aciertos al tratar de detectar 4 emociones. El objetivo de esta investigación era verificar si cabía la posibilidad de detectar las emociones mediante tecnologías como los EEG, lo que quedó claro al leer artículos como el mencionado anteriormente.

A pesar de las esperanzas que proporcionaban estas investigaciones, la utilización de esta tecnología requería de personal con los conocimientos adecuados, con los cuales no contábamos. Sumando a esto la dificultad que añadía la pandemia a la hora de recoger los datos de manera presencial, optamos por tomar un camino en la detección de emociones menos intrusivo.

Como segunda investigación sobre las emociones, teniendo en cuenta nuestro cambio de planteamiento, me encargué de recopilar información sobre las encuestas que se utilizaban en la actualidad para la medición de emociones. Esta investigación me proporcionó un conocimiento sobre SAM (Self-Assessment Manikin), la rueda de Plutchik y las escalas diferenciales, que posteriormente aplicamos sobre nuestros formularios de recogida de datos.

Además de las investigaciones teóricas, también llevamos a cabo una búsqueda de herramientas relacionadas con el tema que queríamos desarrollar. Encontré aplicaciones creadas por el MIT, Microsoft, además de una herramienta llamada EmotionCues. Todas ellas, al observarlas de cerca, contaban con múltiples fallos, se centraban en detectar todo el rango de emociones o funcionaban de manera asíncrona, por lo que no encontré una aplicación en el mercado idéntica a la planteada en un principio. Esta búsqueda nos proporcionó una visión más amplia de la cantidad de herramientas que se ofertan en este sector, y los campos que están por explotar y mejorar.

En la segunda etapa de la investigación, concretamos el tema a desarrollar, enfocándonos en la detección del aburrimiento. En este punto, nos proporcionaron una gran bibliografía que nos repartimos y leímos cada uno. Me centré en un artículo que trataba de predecir el aburrimiento mediante la monitorización del ritmo cardíaco, el nivel de cortisol y la conductividad de la piel. También mencionar el artículo que leí sobre las posturas y gestos faciales que adoptan las personas cuando tienden a estar aburridas, el cual fue de gran utilidad a la hora de identificar los patrones que detectaba la IA final.

Una vez concluida la investigación, comenzamos con la implementación de la IA, de manera colaborativa, para la detección de emociones. Además, en este punto iniciamos la búsqueda de vídeos con licencia CC, de manera que pudiéramos editarlos y mostrarlos legalmente, para la inducción de emociones. Aporté varios vídeos de conferenciantes y profesores, además de uno destacable en el que se mostraba a un hombre tendiendo ropa. Incluí este vídeo sabiendo que era semejante al mencionado en estudios anteriores sobre el aburrimiento, debido a que se había verificado que era habitual que indujera a la persona con este estado.

También se llevó a cabo la búsqueda de una plataforma o método que nos facilitara la recopilación de información y obtención de un mayor dataset. Una de las herramientas por las que optamos fue Amazon mTurk, la cual nos permitía obtener un gran número de respuestas para la clasificación de los vídeos, aunque tuvo que ser descartada debido a su alto coste monetario. Por mi parte, también encontré otras plataformas semejantes a la anterior, pero con menor afluencia de personas debido a su desconocimiento. Debido a todos los inconvenientes, decidimos eliminar esta parte del experimento y comenzar con la etapa del experimento grabado. Contribuí grabando a 4 personas, con la ayuda de Daniel Padilla, visualizando los 19 vídeos acordados, además de pasarles los cuestionarios respectivos al terminar la visualización de cada uno de ellos. Tras obtener todos los datos, y, haciendo uso de unas funciones implementadas de manera colaborativa, transformé las imágenes, el id de la persona y las puntuaciones del vídeo visualizado, en un csv, de manera que se pudiera utilizar junto con el resto de csvs como dataset propio.

En la última etapa de la implementación, modificamos la arquitectura de la IA de manera que nos permitiera detectar el aburrimiento. Para ello, realizamos transfer learning, quedándonos con algunas de las capas ya entrenadas con el dataset de emociones, las cuales ya tendrían ajustados los pesos para esclarecer los patrones que identifican una emoción, y entrenando el resto con nuestro dataset propio, para así poder detectar los patrones que caracterizan el aburrimiento. Sin olvidarnos de la modificación técnica de la arquitectura que llevamos a cabo de manera que la salida variará de 8 clases a 3 (aburrido, neutro, interesante), y la alteración de los parámetros para la obtención de un mejor resultado final.

Durante todo el proceso de elaboración del proyecto, hemos llevado a cabo un registro en la memoria de todo lo investigado e implementado. En mi caso, he realizado un resumen de los papers y herramientas anteriormente mencionadas, junto con algunas otras, en el apartado 2. Además de realizar la redacción de los agradecimientos, la introducción, parte de la reflexión ética y ayudar a mis compañeros en el resto de las secciones.

Artículo Vital Emo: *The boredom detector with a machine learning perspective*

Decidimos escribir un paper ya que consideramos que nuestro proyecto era realmente interesante, y divulgar una investigación de este campo ayudaría a complementar la poca información actual que existe sobre la detección de emociones en los ámbitos de educación online. Presentamos el paper en el *29th International Conference on Computers in Education*.

En la creación de este artículo, me encargué de una primera recopilación de información sobre el aburrimiento, basándome en toda la información que habíamos recopilado desde el comienzo de la investigación del TFG, y su exposición en el apartado Related work, el cual fue posteriormente mejorado y complementado.

Además, contribuí junto con Diego Isar en la recopilación de referencias, garantizando que estuvieran ordenadas con el formato adecuado así como verificar su correcta citación. También llevamos a cabo la traducción y corrección de diversas secciones del artículo.

Índice de figuras

2.1. Escáner de actividad cerebral (Resting State, Boredom, Sustained Attention Interest).	16
2.2. Escáner de actividad cerebral (Boredom, Interest, Resting State).	17
2.3. Escáner de actividad cerebral, perspectiva cenital.	17
2.4. Círculo de emociones de Russell.	21
2.5. Cuestionario pictográfico de Bradley y Lang.	22
2.6. Rueda de las emociones de Plutchik.	23
2.7. Postura de aburrimiento y postura de interés.	24
2.8. Versión resumida del círculo de emociones de Russell.	25
2.9. Guante con sensores.	26
2.10. Ejemplo de funcionamiento de Triplet Lost.	30
2.11. Ejemplo de funcionamiento de Viola Jones.	31
2.12. Ejemplo de funcionamiento de Adaboost.	31
2.13. Ejemplo de funcionamiento de OpenFace.	32
2.14. Modelo VGG16.	33
2.15. Modelo DeepFace Paso 1.	34
2.16. Modelo DeepFace Paso 2.	34
2.17. Arquitectura propuesta por la tesis.	36
2.18. Ejemplo de aprendizaje supervisado.	37
2.19. Ejemplo de aprendizaje no supervisado.	37
2.20. Ejemplo de aprendizaje por refuerzo.	37
2.21. Jerarquía de conceptos respecto a la Inteligencia Artificial.	38
2.22. Ilustración del funcionamiento conceptual de una red neuronal.	38
2.23. Ilustración del funcionamiento lógico de una neurona artificial.	39
2.24. Función sigmoide.	39
2.25. Función de tangente hiperbólica.	40
2.26. Función ReLU.	40
2.27. Función Softmax.	41
2.28. Ejemplo de capas en las redes neuronales convolucionales.	41
2.29. Ejemplo de una imagen a nivel técnico.	42
2.30. Ejemplo del procedimiento de las convoluciones básicas.	42
2.31. Ejemplo del procedimiento de las convoluciones con stride = 3.	43
2.32. Ejemplo de distintas implementaciones de Pooling.	43
2.33. Ejemplo de vectorización de una imagen.	44
2.34. Ejemplo de red estándar (arriba) y red recurrente a corto plazo (abajo).	44
2.35. Red Neuronal Recurrente a Largo Plazo (LSTM).	45
2.36. Funcionalidades de iMotions.	47
2.37. Ejemplo de uso de iMotions.	47

2.38. Ejemplo de uso de EmotionCues.	48
2.39. Búsqueda de rostros con el uso de Face API.	48
2.40. Predicciones de los rostros analizados por Face API.	49
2.41. Sensor diseñado por el MIT.	50
3.1. Planificación del proyecto.	53
3.2. Objetivos del proyecto.	53
4.1. Ejemplo de experimento simulado.	55
4.2. Trabajos ofrecidos para Turkers en Amazon MTurk.	56
4.3. Diseño experimental de Amazon MTurk.	59
5.1. Patrón de capas utilizado	60
5.2. Arquitectura del proyecto.	61
5.3. Imagen anómala detectada.	63
5.4. Línea anómala.	63
6.1. Resultado de las puntuaciones de los vídeos.	64
6.2. Precisión del modelo de las emociones.	65
6.3. Coste del modelo de las emociones.	65
6.4. Precisión del modelo del aburrimiento.	66
6.5. Coste del modelo del aburrimiento.	66
6.6. Detección facial y su porcentaje de aburrimiento.	67
6.7. Ejemplo de uso real en una reunión.	67
9.1. Ejemplo de clasificación con el porcentaje.	101

Índice de cuadros

2.1. Emociones opuestas.	18
2.2. Escala diferencial.	20
4.1. Encuesta de datos personales.	56
4.2. Encuesta de calificación de vídeos.	56
4.3. Ejemplo de clasificación.	57
4.4. Casillas de la encuesta en Amazon MTurk.	57
4.5. Distribución de pesos Amazon MTurk.	57
4.6. Diseño experimental de Amazon MTurk.	58
6.1. Distribución de pesos Amazon MTurk.	67
6.2. distribución usada.	68

Bibliografía

- Ahonen, T., Hadid, A., and Pietikäinen, M. (2004). Face recognition with local binary patterns. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- AIED/ITS (2021). <https://icce2021.apsce.net/call-for-papers/c1/>.
- Alandete, D. (2011). John McCarthy, el arranque de la inteligencia artificial. <https://elpais.com/diario/2011/10/27/necrologicas/1319666402{ }850215.html>.
- ALEKS (2020). ALEKS – Adaptive Learning Assessment for Math, Chemistry, Statistics More. <https://www.aleks.com/>.
- Amazon (2005). Amazon MTurk.
- Amigo, N. (2014). El Ibex está... aburrido y pesado. <https://www.youtube.com/watch?v=0H9hGVZY74w>.
- Amos, B., Ludwiczuk, B., and Satyanarayanan, M. (2016). OpenFace: A General-Purpose Face Recognition Library with Mobile Applications. <http://cmusatyalab.github.io/openface/>.
- Andrews-Hanna, J. R. (2012). The brain’s default network and its adaptive role in internal mentation.
- Attention Research (2016). Borapp Bzzz. <https://apkpure.com/es/borapp-bzzz/org.pielot.borapp3>.
- Bach, D. R., Daunizeau, J., Friston, K. J., and Dolan, R. J. (2017). PsPM-SCRV4: Skin conductance responses in fear conditioning with visual CS and auditory US.
- Badiou, A. (2012). La República de Platón. <https://www.youtube.com/watch?v=SihN45cBL8k>.
- Baritz, M. I. (2016). Analysis of human bio-behavior and eye movements correlated with induced positive/negative emotional level. In *Conference proceedings of» eLearning and Software for Education «(eLSE)*, number 01, pages 38–44. ” Carol I” National Defence University Publishing House.
- BBC (2006). Emotion sensor ’detects boredom’. <http://news.bbc.co.uk/2/hi/health/4856050.stm>.
- Beck, A. T., Steer, R. A., and Brown, G. (1996). Beck depression inventory–ii. *Psychological Assessment*.
- Berlyne, D. E. (1960). Conflict, arousal, and curiosity.

- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.
- Britch, D. and Olprod (2018). Reconocimiento de emociones percibido mediante el Face API. <https://docs.microsoft.com/es-es/xamarin/xamarin-forms/data-cloud/azure-cognitive-services/emotion-recognition>.
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain’s default network: anatomy, function, and relevance to disease.
- Cabrera, L., Benítez, J. T. B., Afonso, M. G., and Pérez, P. Á. (2006). Un estudio transversal retrospectivo sobre prolongación y abandono de estudios universitarios. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 12(1).
- Carrier, P.-L. and Courville, A. (2013). Challenges in representation learning: Facial expression recognition challenge. <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>.
- Carriere, J. S., Cheyne, J. A., and Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and cognition*, 17(3):835–847.
- Challenge, K. F. (2018). FER Challenge. <https://www.kaggle.com/ashishpatel26/facial-expression-recognitionferchallenge>.
- Chang, C. I., Wang, Y., and Chen, S. Y. (2015). Anomaly Detection Using Causal Sliding Windows. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Cheyne, J. A., Carriere, J. S., and Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and cognition*, 15(3):578–592.
- Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences of the United States of America*.
- Cohen, S., Doyle, W. J., Skoner, D. P., Fireman, P., Gwaltney, J. M., and Newsom, J. T. (1995). State and Trait Negative Affect as Predictors of Objective and Subjective Symptoms of Respiratory Viral Infections. *Journal of Personality and Social Psychology*, 68(1):159–169.
- Commons, C. (2001). Creative Commons. <https://creativecommons.org/>.
- Communications, B. (2021). ¿Quiénes son los ‘millennials’ y por qué son una generación única? <https://www.bbva.com/es/quienes-millennials-generacion-unica/>.
- Cowen, A. S. and Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, 114(38):E7900.
- Dalby, M. (2013). Hanging washing. <https://www.youtube.com/watch?v=pXXN10cyxZQ>.

- Danckert, J. and Isacescu, J. (2017). The bored brain : insular cortex and the default mode network . pages 1–15.
- Danckert, J. and Merrifield, C. (2018). Boredom, sustained attention and the default mode network. *Experimental Brain Research*.
- de la Torre, D. (2019). ¿Cómo nació la Inteligencia Artificial? <https://blogthinkbig.com/historia-como-nacio-inteligencia-artificial>.
- ElBoletín (2016). Apple compra una ‘start up’ capaz de reconocer las emociones.
- Emotient (2012). Emotient. <http://emotient.com>.
- Europea, C. (2021). Ciberseguridad Europea. <https://www.consilium.europa.eu/es/policies/cybersecurity/>.
- European Commission (2016). TeSLA. <http://tesla-project-eu.azurewebsites.net>.
- Farmer, R. and Sundberg, N. D. (1986). Boredom proneness—the development and correlates of a new scale. *Journal of Personality Assessment*, 50(1):4–17. PMID: 3723312.
- Fortuna, C. (2014). ¿Cómo piensa un mago? <https://www.youtube.com/watch?v=kQW8VsLrnXw>.
- Gao, W. and Lin, W. (2012). Frontal Parietal Control Network Regulates theAnti- Correlated Default and Dorsal AttentionNetworks.
- Gonzalez Pacheco, V. (2019). Una Breve Historia del Machine Learning. <https://empresas.blogthinkbig.com/una-breve-historia-del-machine-learning/>.
- Gooru (2021). Navigator. <https://gooru.org/about/>.
- Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*.
- Guo, Q., Feng, W., Zhou, C., Huang, R., Wan, L., and Wang, S. (2017). Learning Dynamic Siamese Network for Visual Object Tracking. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial pre-frontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*.
- Hochrangige Expertengruppe für Künstliche Intelligenz (HEG-KI) (2020). *Assessment List for Trustworthy AI (ALTAI)*.
- Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*.
- ICCE (2021). ICCE 2021 Toc. <https://icce2021.apsce.net/>.
- IEEE (2021). IEEE. <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>.
- IMotions (2006). iMotions. <https://imotions.com>.

- ISO (2017). ISO. <https://www.iso.org/committee/6794475.html>.
- Kim, J., Seo, J., and Laine, T. H. (2018). Detecting boredom from eye gaze and EEG. *Biomedical Signal Processing and Control*, 46:302–313.
- Kim, M., Kumar, S., Pavlovic, V., and Rowley, H. (2008). Face tracking and recognition with visual constraints in real-world videos. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*.
- Knapp, P. H., Levy, E. M., Giorgi, R. G., Black, P. H., Fox, B. H., and Heeren, T. C. (1992). Short-term immunological effects of induced emotion. *Psychosomatic Medicine*, 54(2):133–148.
- Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). DEAP: A database for emotion analysis; Using physiological signals. *IEEE Transactions on Affective Computing*.
- Kroes, S. (2007). Detection Boredom in Meetings. pages 3–7.
- Kumari, R. and Kr., S. (2017). Machine Learning: A Review on Binary Classification. *International Journal of Computer Applications*.
- La, P. (2021). S1ronda de enne a 10 de ung. page 2021.
- Lacey, J. I. (1959). Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In *Research in Psychotherapy, Apr, 1958, Washington, DC; This conference, financed by a grant (M-2031) from the National Institute of Mental Health, US Public Health Service, was held under the auspices of the Division of Clinical Psychology, American Psychological Association, with planning and programming by an Ad Hoc Committee of the Division of Clinical Psychology; Frank Auld, Jr., Morris B. Parloff, Benjamin Pasamanick, George Saslow, Julius Seeman, and Eli A. Rubinstein, Chairman*. American Psychological Association.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*.
- Lee, Y. Y. and Hsieh, S. (2014). Classifying different emotional states by means of eeg-based functional connectivity patterns. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095415>.
- Lin, Y., Shen, J., Cheng, S., and Pantic, M. (2020). FT-RCNN: Real-Time Visual Face Tracking with Region-Based Convolutional Neural Networks. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG)*, pages 267–274.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C. (2016). SSD: Single shot multibox detector. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Liu, Y. (2020). Consistent and Accurate Face Tracking and Recognition in Videos. http://rave.ohiolink.edu/etdc/view?acc{_}num=ohiou1588598739996101.
- Lons, P. (2021). Mentiras que te dicen en la Clase de Historia. <https://www.youtube.com/watch?v=627d1CJJTDI>.

- M. Gopi Krishna, A. Srinivasulu, and Prof (Dr.) T.K.Basak (2012). Face Detection System on Ada boost Algorithm Using Haar Classifiers. *International Journal of Modern Engineering Research (IJMER)*, Volume 2(Issue 6):3996–4000.
- Malkovsky, E., Merrifield, C., Goldberg, Y., and Danckert, J. (2012). Exploring the relationship between boredom and sustained attention. *Experimental Brain Research*, 221(1):59–67.
- Mangahigh (2008). Mangahigh. <https://www.mangahigh.com/es-es/>.
- Marchesi, A. and Hernandez Gil, C. (2003). El fracaso escolar: una perspectiva internacional. *El fracaso escolar: una perspectiva internacional, 2003, ISBN 84-206-2955-3, págs. 319-329*, pages 319–329.
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810):393–395.
- Matheson, R. (2017). Wearable device reveals consumer emotions. *MIT News*.
- Mehrabian, A. et al. (1971). *Silent messages*, volume 8. Wadsworth Belmont, CA.
- Meirieu, P. (2019). ¿Qué Pedagogía para este tiempo? <https://www.youtube.com/watch?v=FRuGKYV2NTY>.
- Merrifield, C. and Danckert, J. (2014). Characterizing the psychophysiological signature of boredom. *Experimental Brain Research*, 232(2):481–491.
- Microsoft (2010). Face. <https://azure.microsoft.com/es-es/services/cognitive-services/face/#demo>.
- Mikulas, W. and Vodanovich, S. (1993). *The essence of boredom. The Psychological Record*.
- Mishra, M. (2020). Convolutional Neural Networks, Explained. <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>.
- Ng, H. W., Nguyen, V. D., Vonikakis, V., and Winkler, S. (2015). Deep learning for emotion recognition on small datasets using transfer learning. In *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, pages 443–449. Association for Computing Machinery, Inc.
- Ng, H. W., Wang, X., Gao, Y., and Ma, R. (2020). ENHANCED FACE-DETECTION AND FACE-TRACKING FOR EMBEDDED VISION SYSTEMS.
- Nielek, R., Ciastek, M., and Kopeć, W. (2017). Emotions make cities live. Towards mapping emotions of older adults on urban space. *Proceedings - 2017 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2017*, (September):1076–1079.
- OpenCV (1999). OpenCV. <https://opencv.org><https://github.com/opencv/opencv>.
- Oppermann, A. (2019). What is Deep Learning and How does it work? <https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac>.

- Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The measurement of meaning*. Number 47. University of Illinois press.
- O'Shea, K. and Nash, R. (2015). An Introduction to Convolutional Neural Networks. (December).
- Paniagua, J. C. (2017). Quantum theory (1st part): so precise, so astonishing. <https://www.youtube.com/watch?v=N-w1tkvdsQI>.
- Paulus, P. C., Castegnetti, G., and Bach, D. R. (2019). PsPM-HRM_IAPS: SCR, ECG and respiration measurement in response to aversive/arousing/neutral IAPS pictures.
- Pearson (2016). WriteToLearn. A Web-Based AI-Automated Writing Scoring, Feedback, and Reporting Platform for Students. <https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/WriteToLearn/p/100000030.html?tab=faqs>.
- Pekrun, R. (2005). Progress and open problems in educational emotion research. *Learning and Instruction*, 15(5):497–506.
- Pena, F. (2018). Cómo seducir a tu cliente. <https://www.youtube.com/watch?v=m06sgGAEjp0>.
- Pena, F. (2019). La ciencia de la influencia y la persuasión. https://www.youtube.com/watch?v=7B5CrH2jeWg&ab_channel=FernandoPena.
- Phi, M. (2018). Illustrated Guide to Recurrent Neural Networks. <https://towardsdatascience.com/illustrated-guide-to-recurrent-neural-networks-79e5eb8049c9>.
- Plawiak, P., Sośnicki, T., Niedźwiecki, M., Tabor, Z., and Rzecki, K. (2016). Hand Body Language Gesture Recognition Based on Signals From Specialized Glove and Machine Learning Algorithms. *IEEE Transactions on Industrial Informatics*, 12(3):1104–1113.
- Plutchik, R. (1991). *The emotions*. University Press of America.
- Portuondo, M. M. (2019). La primera vuelta al mundo. <https://www.youtube.com/watch?v=XrrnrU-K9o>.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*.
- Rebollo-Catalán, Á., Hornillo Gómez, I., and García Pérez, R. (2006). El estudio educativo de las emociones: una aproximación sociocultural. *Teoría de la Educación - Educación y Cultura en la Sociedad de la Información*, 7(2):28–44.
- Rebollo Catalán, M. Á., García Pérez, R., Barragán Sánchez, R., Buzón García, O., and Vega Caro, L. (2008). Las emociones en el aprendizaje online. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 14(1).
- Recuero de los Santos, P. (2017). Tipos de aprendizaje en Machine Learning: supervisado y no supervisado.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified,

- real-time object detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Reeve, J. and Nix, G. (1997). Expressing intrinsic motivation through acts of exploration and facial displays of interest. *Motivation and Emotion*, 21(3):237–250.
- Rosnow, R. L. (2004). Semantic differential.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.
- S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. P. (2012). DEAP: A Database for Emotion Analysis using Physiological Signals.
- Samper, M. B. (2020). Reglamento (Ue) 2016/679 Del Parlamento Europeo Y Del Consejo De 27 De Abril De 2016, Relativo a La Protección De Las Personas Físicas En Lo Que Respecta Al Tratamiento De Datos Personales Y a La Libre Circulación De Estos Datos Y Por El Que Se Deroga La . *Protección de datos personales*, 2014:17–144.
- Sanabria-Macías, F., Romera, M. M., Macías-Guarasa, J., Pizarro, D., Turnes, J. N., and Reyes, E. J. M. (2019). Face tracking with a probabilistic Viola and Jones face detector. In *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 5616–5621. IEEE.
- Sarkar, D. (2018). A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning. <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep->
- Schmitt, M. (2020). What Is Machine Learning? — A visual explanation. <https://towardsdatascience.com/what-is-machine-learning-a-visual-explanation-14642b90429f>.
- Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823.
- Shan, C., Gong, S., and McOwan, P. W. (2009). Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816.
- Sharma, B. and Bhatia, R. (2020). Development of an improved Face Recognition Detection and Classification Method.
- Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, 5(3):270–290.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14.
- Solomon, R. C. and Stone, L. D. (2002). On “ Positive ” and “ Negative ” Emotions. pages 417–435.

- Sotelo, L. (2016). Introducción a la Escuela de Frankfurt. <https://www.youtube.com/watch?v=WWYbanKdyeA>.
- Stevenson, R. A. and James, T. W. (2008). Affective auditory stimuli: Characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. *Behavior Research Methods*.
- Struk, A. A., Scholer, A. A., and Danckert, J. (2016). A self-regulatory approach to understanding boredom proneness. *Cognition and Emotion*, 30(8):1388–1401.
- Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708.
- Thinkster (2015). Only Math Tutoring Program. <https://hellothinkster.com#>.
- Tomás Abraham, D. P. (2015). Historia de la Filosofía. <https://www.youtube.com/watch?v=RXOfas8YxNQ>.
- Toor, A. (2017). This French school is using facial recognition to find out when students aren't paying attention. *The Verge*.
- Unicheck (2017). Unicheck. <https://unicheck.com>.
- Vodanovich, S. J., Wallace, J. C., and Kass, S. J. (2005). A confirmatory approach to the factor structure of the boredom proneness scale: Evidence for a two-factor short form. *Journal of Personality Assessment*, 85(3):295–303. PMID: 16318568.
- Wang, Y.-Q. (2014). An Analysis of the Viola-Jones Face Detection Algorithm. *Image Processing On Line*, 4:128–148.
- Zamanian, H. and Farsi, H. (2018). A new feature extraction method to improve emotion detection using EEG signals. *Electronic Letters on Computer Vision and Image Analysis*, 17(1):29–44.
- Zeng, H., Shu, X., Wang, Y., Wang, Y., Zhang, L., Pong, T.-C., and Qu, H. (2020). EmotionCues : Emotion-Oriented Visual Summarization of Classroom Videos . *IEEE Transactions on Visualization and Computer Graphics*, 27(7):3168–3181.
- Zhang, Q., Chen, X., Zhan, Q., Yang, T., and Xia, S. (2017a). Respiration-based emotion recognition with deep learning. *Computers in Industry*, 92-93:84–90.
- Zhang, Q., Zhang, M., Chen, T., Sun, Z., Ma, Y., and Yu, B. (2019). Recent advances in convolutional neural network acceleration. *Neurocomputing*, 323:37–51.
- Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., and Li, S. Z. (2017b). S3FD: Single Shot Scale-Invariant Face Detector. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1):43–76.

Anexos

Anexo 1 - Ética

Este Trabajo de Fin de Grado tiene como objetivo ayudar a los docentes a impartir clases y evitar que los alumnos desconecten, la herramienta implementada debería ser la base para un futuro proyecto que permita a los docentes conocer el estado de desconexión del alumnado y así, cambiar la metodología de impartir clase en tiempo real.

Debido a que el desarrollo de este proyecto es, en gran parte, la creación de una Inteligencia Artificial, se ha visto la obligación de hacer un análisis retrospectivo sobre la fiabilidad del producto final así como los usos que se le pueden llegar a dar.

Las características que debe cumplir una IA para ser fiable, según Hochrangige Expertengruppe für Künstliche Intelligenz (HEG-KI) (2020) son:

- **Debe ser lícita.** La IA debe cumplir la carta de Derechos fundamentales así como las leyes y reglamentos aplicables.
- **Ha de ser ética.** Debido a que las leyes avanzan a un ritmo más lento que la evolución tecnológica pueden darse normas o leyes que son incoherentes o poco adecuadas frente a las normas éticas. La IA debería garantizar el respeto de los principios y valores éticos.
- **Deben ser robustas tanto desde un punto de vista técnico como social.** Aunque las intenciones del uso sean buenas se pueden provocar daños accidentales.

Estos puntos son necesarios pero no suficientes como para asegurar la fiabilidad de una IA por lo que a continuación se exponen unos requisitos y cuestiones aplicados al proyecto.

9.4.1. Capacidad humana de decisión y Supervisión

Capacidad humana de decisión

- **¿El Sistema de Inteligencia Artificial está diseñado para interactuar, guiar o tomar decisiones por parte de usuarios finales humanos que afecte a otros humanos o la sociedad?**

Sí, ya que el sistema está diseñado para guiar a un docente o conferenciante a través de su exposición. Esto sirve únicamente como una ayuda para predecir cuándo cambiar el ritmo de la clase, realizar preguntas o poner ejemplos que capten la atención del público.

- **¿Podría el sistema de IA generar confusión para algunos o todos los usuarios finales sobre si una decisión, contenido, consejo o resultado es el resultado de una decisión algorítmica?**

No aplica, actualmente nuestra Inteligencia Artificial no aconseja, en todo caso tampoco consideramos que pueda generar confusión, ya que en el supuesto de seguir con el proyecto es probable que lo anunciemos como una inteligencia artificial y por lo tanto de una información basada en algoritmos.

- **¿Se informa adecuadamente a los usuarios finales de que una decisión, contenido, consejo o resultado es el resultado de una decisión algorítmica?**

Sí, ya que los estudios pueden determinar si una persona está aburrida o interesada, pero nuestra inteligencia artificial muestra un ‘porcentaje de seguridad’ respecto al acierto que tiene, por ejemplo, un porcentaje al 90 % indica que el algoritmo está seguro al 90 %.

- **¿Podría el sistema de Inteligencia Artificial generar confusión a los usuarios finales que están interactuando con ella sobre si se están comunicando con un humano o una IA?**

No. En todo momento se puede apreciar el hecho de que se está tratando con una IA. El usuario trata directamente con una imagen del oyente en la que se puede observar el estado predicho y el porcentaje de seguridad.

- **¿Se informa a los usuarios finales de que están interactuando con un sistema de IA?**

Sí, actualmente se informa de manera verbal que se está interactuando con un sistema de inteligencia artificial.

- **¿Puede el sistema de Inteligencia Artificial afectar a la autonomía humana creando una dependencia de su uso?**

No, esta inteligencia artificial nunca se pensó como sustitutiva de un ser humano, sino como apoyo para mejorar las clases online. En el caso de que sea sustitutivo o se cree dependencia, recomendamos encarecidamente que se prescinda de esta herramienta.

- **¿Implementó procedimientos para evitar que los usuarios finales confíen demasiado en el sistema de inteligencia artificial?**

Sí, el porcentaje de certeza nunca llegará al 100 % porque la herramienta nunca podría afirmar rotundamente que su predicción es correcta, por lo que siempre muestran porcentajes y siempre son variables, es decir, que se tienen que tomar como información o referencia pero nunca como certeza.

- **¿Puede la Inteligencia Artificial interferir con la voluntad del usuario final y actuar de una forma no deseada?**

No. Todo el efecto que tiene el uso de esta Inteligencia Artificial es mostrar el estado de una persona y el nivel de seguridad. En ningún momento participa de

forma activa en la interacción entre dos o más personas si no que tiene un uso pasivo e informativo.

- **¿Existe algún procedimiento para evitar que el sistema de IA afecte inadvertidamente la autonomía humana?**

No aplica ya que no interferimos con la autonomía humana.

- **¿El sistema de IA simula la interacción social con o entre usuarios finales o sujetos?**

No aplica ya que no simula la interacción.

- **¿El sistema de IA corre el riesgo de crear apego humano, estimular el comportamiento adictivo o manipular el comportamiento del usuario?**

No aplica ya que no simula, no es adictiva y no busca manipular el comportamiento del usuario, en todo caso busca informar al docente para que este cambie conscientemente su forma de dar clase.

Supervisión

- **Por favor, determine si el sistema de IA (elija tantos como sea apropiado):**

- **Es un sistema de autoaprendizaje o autónomo**

No.

- **Es supervisado por un humano dentro del bucle (HITL)**

No.

- **Es supervisado por un humano fuera del bucle (HOTL)**

No.

- **Es supervisado por un humano al mando (HIC)**

Sí, ya que el docente decide dependiendo de la situación que le muestre la herramienta.

- **¿Las personas que hacen uso de esta herramienta han sido formadas para realizar la supervisión?**

La idea principal es que, antes de usarse en un caso real, se enseñe y explique al docente el funcionamiento de la herramienta además de informarle del porcentaje de precisión que tiene en todo momento o si se han detectado sesgos. El docente debe conocer cómo debe estar situada la cámara de la audiencia para que no surjan falsos positivos.

- **¿Existe algún mecanismo de detección y respuesta para los efectos adversos indeseables del sistema de IA para el usuario final o el sujeto?**

No aplica, nuestra herramienta es informativa.

- **¿Se ha asegurado un “botón de stop” o procedimiento para abortar el funcionamiento cuando sea necesario?**

Sí, en cualquier momento el docente puede presionar una tecla concreta para detener la ejecución del programa.

- **¿Existe alguna medida específica de supervisión y control para reflejar el autoaprendizaje o la naturaleza autónoma del sistema de IA?**

No aplica, nuestra herramienta no es autónoma, es informativa

9.4.2. Robustez técnica y seguridad

Robustez técnica y seguridad

- **¿Podría el sistema de IA tener efectos adversos, críticos o dañinos (por ejemplo, para la seguridad humana o social) en caso de riesgos o amenazas como fallas técnicas o de diseño, defectos, interrupciones, ataques, mal uso, uso inapropiado o malicioso?**

No, la herramienta se ha planteado con un único objetivo informativo para ayudar en la labor docente. En ningún caso se debe tomar el resultado obtenido como verdad absoluta, pudiendo ignorarlo en cualquier momento.

- **¿El sistema de IA está certificado para ciberseguridad (por ejemplo, el esquema de certificación creado por la Ley de Ciberseguridad Europea (2021)) o cumple con estándares de seguridad específicos?**

No, la herramienta actualmente no podría considerarse ni en versión pre-alpha, por lo que no cumple estándares concretos.

- **¿Qué tan expuesto está el sistema de IA a los ciberataques?**

No aplica, nuestra herramienta actualmente es local y su uso es en un entorno controlado. El grupo es consciente de que, si el proyecto sale adelante, sería lógico invertir en seguridad, ya que su uso sería mediante una API, página web o programa con conexión a internet y sí, la aplicación estaría tan expuesta como cualquier otro servicio de internet.

- **¿Implementó medidas para garantizar la integridad, solidez y seguridad general del sistema de inteligencia artificial contra posibles ataques durante su ciclo de vida?**

No aplica, actualmente nuestra herramienta es un prototipo y nos hemos centrado exclusivamente en la funcionalidad de la misma.

- **¿Se ha hecho algún test de penetración?**

No aplica, actualmente nuestra herramienta es un prototipo.

- **¿Informó a los usuarios finales sobre la duración de la cobertura de seguridad y las actualizaciones?**

No aplica, actualmente nuestra herramienta es un prototipo y su uso es en un entorno controlado.

Seguridad general

- **¿Se han definido los riesgos, sus métricas y sus niveles en la IA en cada caso de uso específico?**

No se han definido, el único riesgo que puede conllevar es un estrés añadido al docente, ya que puede no saber cómo afrontar la situación en el caso de aburrir al alumnado.

- **¿Implementó un proceso para medir y evaluar los riesgos de manera continua?**

No aplica, ya que nuestra herramienta es un prototipo, aún así consideramos que sería interesante implementar un formulario de contacto con un apartado de riesgos en el que los propios alumnos o docentes puedan informarnos de situaciones de riesgo.

- **¿Informó a los usuarios finales y a los sujetos de los riesgos existentes o potenciales?**

No aplica, ya que nuestra herramienta es un prototipo, aún así consideramos que sería interesante implementar un informe de los riesgos existentes y potenciales de la herramienta, por ejemplo, el mal uso de la misma, llevando al extremo su uso y llegando a frustrar al docente.

- **¿Se han identificado fallos de diseño, fallos técnicos, amenazas ambientales y sus posibles consecuencias?**

Sí, ya que nuestra herramienta es un prototipo, no cuenta con una arquitectura final pulida y un porcentaje alto de aciertos. Somos conscientes de que si continuamos con el desarrollo de esta herramienta, debemos eliminar estos fallos para así poder introducirla en el ámbito docente.

- **¿Se ha evaluado el riesgo de un posible uso malintencionado o inadecuado del sistema de inteligencia artificial?** Sí, se podría llegar a dar la frustración del docente a causa de no alcanzar el 100 % de interés entre sus alumnos.

- **¿Se han definido los niveles de criticidad de la seguridad (por ejemplo, relacionados con la integridad humana) de las posibles consecuencias de fallos o mal uso del sistema de IA?** No aplica, nuestra herramienta es informativa, como hemos comentado anteriormente, sí, es posible que se pueda ‘atacar’ a la integridad de un docente en el caso de que este haga un uso de la herramienta llevándola hasta el extremo y buscando un porcentaje de interés perfecto. Por otro lado, desde el punto de vista del alumnado, este es libre de encender o no la cámara, y deberá ser informado de que se van a estudiar sus patrones faciales, por lo que solo podría tener un ‘nivel crítico de seguridad’ (se le estudia la cara) en el caso de asumirlo.

- **¿Se ha evaluado la dependencia de las decisiones críticas del sistema de IA en su comportamiento estable y confiable?**

No, actualmente es un prototipo y no garantiza la obtención de un resultado confiable. Si se continúa con su desarrollo, somos conscientes de que este aspecto debe cambiar para así poder incorporarlo de manera segura en ámbitos educativos.

- **¿Se ha previsto la tolerancia a los fallos mediante, por ejemplo, un sistema duplicado u otro sistema paralelo (basado en IA o convencional)?**

No, ya que es un prototipo, no se ha llevado a cabo un estudio de la tolerancia a los fallos de nuestra herramienta. En las siguientes etapas de su desarrollo sería conveniente realizar pruebas de las diferentes partes de la herramienta, una vez consigamos la arquitectura y datasets óptimos.

- **¿Se ha desarrollado un mecanismo para evaluar cuándo el sistema de IA ha sido modificado para merecer una nueva revisión de su solidez técnica y seguridad?**

No, actualmente la herramienta está en una fase de prototipo, por lo que todas las modificaciones realizadas han tenido como objetivo crear la estructura que garantizara su funcionamiento. En las siguientes etapas de su desarrollo llevaremos a cabo este tipo de pruebas cuando las modificaciones afecten a diversas secciones de la herramienta o añadan grandes cambios.

Precisión

- **¿Un bajo nivel de precisión del sistema de IA podría tener consecuencias críticas, adversas o perjudiciales?**

El diseño de la herramienta se ha realizado con el objetivo de informar al docente. En el caso de obtener un bajo nivel de precisión, podría degenerar en una frustración del docente hacia su método de impartir clase o hacia la herramienta. Debido a esto, informamos en todo momento de la fase prototipo en la que se encuentra la herramienta y advertimos de no tomar sus resultados como verdad absoluta. Para garantizar la confianza de los docentes e incorporar su uso en las aulas debemos alcanzar un alto porcentaje de precisión, asegurando así su utilidad.

- **¿Se han puesto en marcha medidas para garantizar que los datos (incluidos los de entrenamiento) utilizados para desarrollar el sistema de IA están actualizados, son de alta calidad, están completos y son representativos del entorno en el que se desplegará el sistema?**

Los datos están actualizados, son de una calidad media, ya que el propio usuario puntuaba lo que sentía en ese momento y, quizás tras algunos vídeos, no se rellena con números precisos la encuesta, los datos están completos y no son completamente representativos respecto al entorno ya que algunas personas que se han grabado no han sido alumnos.

- **¿Se han puesto en marcha una serie de medidas para controlar y documentar la precisión del sistema de IA?**

Actualmente la única documentación de la precisión de esta herramienta es esta memoria y el propio paper.

- **¿Se ha considerado si el funcionamiento del sistema de IA puede invalidar los datos o las suposiciones con las que se ha entrenado, y cómo esto podría llevar a efectos adversos?**

La inteligencia artificial no debería entrenarse en el momento en el que se está informando al docente ya que no hemos programado un aprendizaje continuo, y por

lo tanto no sería posible invalidar los datos, ya que el entrenamiento lo haríamos nosotros y luego liberaríamos las versiones correspondientes.

En todo caso, la invalidación de datos o suposiciones posibles podría ser que la inteligencia artificial dejase de detectar patrones relacionados con el interés y el aburrimiento y cambiarlos por otros. Los efectos adversos que tendría son la desconexión de la clase por el cambio de metodología del docente, ya que, por ejemplo, puede que los alumnos estén completamente interesados en la clase y que una modificación de los parámetros informase al docente de que la clase es aburrida, cuando esto no es así.

- **¿Se han puesto en marcha procesos para garantizar que se comuniquen adecuadamente el nivel de precisión del sistema de IA que deben esperar los usuarios finales y/o los sujetos?**

Sí, en cada predicción realizada mostramos la clasificación de la emoción (Aburrido, Neutro, Interesado) junto con su porcentaje de precisión.

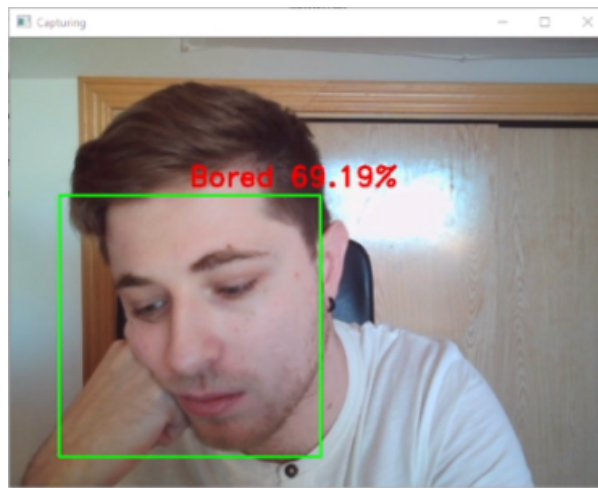


Figura 9.1: Ejemplo de clasificación con el porcentaje.

Fiabilidad, planes alternativos y reproducción

- **¿Podría el sistema de IA causar consecuencias críticas, adversas o perjudiciales (por ejemplo, relativas a la seguridad humana) en caso de baja fiabilidad y/o reproducibilidad?**

No, nuestra herramienta tiene como único objetivo informar al docente. En caso de baja fiabilidad, el docente podrá optar por ignorar el resultado recibido y continuar con su método de impartir clase.

- **¿Se ha definido un proceso para controlar si el sistema de IA cumple los objetivos previstos?**

Al terminar con la implementación de la herramienta, llevamos a cabo un proceso de verificación del cumplimiento de los objetivos planteados, recopilado en el apartado Discusión de resultados de la memoria. Respecto al control de los objetivos una vez se implemente la herramienta en el ámbito docente, aún

está por definir, debido a que al ser un prototipo, no está preparada para realizar pruebas en entornos no controlados.

- **¿Se ha comprobado si hay que tener en cuenta contextos o condiciones específicas para garantizar la reproducibilidad?**

Si, nuestra herramienta puede ser utilizada en cualquier condición o contexto, aunque obtiene un menor número de falsos positivos al tener una mejor iluminación.

- **¿Se han puesto en marcha métodos de verificación, validación y documentación (por ejemplo, registros) para evaluar y garantizar diferentes aspectos de la fiabilidad y reproducibilidad del sistema de IA?**

No aplica, al ser un prototipo, no hemos podido desarrollar actividades de verificación y validación en el entorno para el que está destinado. Respecto a la documentación, se ha desarrollado una recopilación del proceso de construcción y obtención de resultados en la memoria.

- **¿Se han documentado y puesto en marcha los procesos de prueba y verificación de la fiabilidad y reproducibilidad del sistema de IA?**

No aplica, al ser un prototipo, no hemos podido desarrollar pruebas de verificación, fiabilidad y reproducibilidad en el entorno para el que está destinado.

- **¿Se han definido planes de seguridad a prueba de fallos para hacer frente a los errores del sistema de IA, sea cual sea su origen, y se han puesto en marcha procedimientos de gobernanza para activarlos?**

No, al ser un prototipo, somos conscientes de que cuenta con fallos que deben ser arreglados. En futuras líneas de desarrollo, una vez tengamos una arquitectura y estructura consolidada, deberemos elaborar planes de seguridad advirtiendo al docente de los fallos o baja precisión.

- **¿Se ha establecido un procedimiento adecuado para tratar los casos en los que el sistema de IA ofrezca resultados con una puntuación baja?**

En todo momento nuestra herramienta es informativa y comunicamos al docente el nivel de precisión con el que la IA predice el resultado. En caso de obtener una baja puntuación, el docente puede decidir ignorar el resultado obtenido y continuar con su exposición.

- **¿El sistema de IA utiliza el aprendizaje continuo (en línea)?**

No, actualmente la IA solo se actualiza en el momento que nosotros decidamos volver a entrenarla con nuevos datos, no adquiere conocimiento por otros medios.

- **¿Has tenido en cuenta las posibles consecuencias negativas de que el sistema de IA aprenda métodos novedosos o inusuales para obtener una buena puntuación en su función objetivo?**

No, ya que no hemos implementado la herramienta para que aprenda de manera continua.

9.4.3. Privacidad y Data Governance

Privacidad

- **¿Se ha considerado el impacto del sistema de IA en el derecho a la intimidad, el derecho a la integridad física, mental y/o moral y el derecho a la protección de datos?**

Si, hemos desarrollado la herramienta tratando de ser lo menos intrusivos posible, sin almacenar nuevos datos personales y avisando previamente del tratamiento que se va a realizar con estos. Todo alumno puede negarse al tratamiento de sus datos y por lo tanto no encender su cámara.

- **En función del caso de uso, ¿se han establecido mecanismos que permitan señalar cuestiones relacionadas con la privacidad en relación con el sistema de IA?**

No, actualmente contamos con una herramienta en fase prototipo que solo se ha puesto en marcha en situaciones controladas. En un futuro desarrollo sería interesante crear un foro o email donde los usuarios nos hicieran llegar sus preocupaciones, para así solucionar los posibles problemas de privacidad que pudiera generar nuestra herramienta.

Data Governance

- **¿El sistema de IA está siendo entrenado o ha sido desarrollado, utilizando o procesando datos personales (incluyendo categorías especiales de datos personales)?**

Si, la herramienta ha sido entrenada utilizando los rostros de las personas que participaron y dieron su consentimiento en el experimento de recopilación de datos.

- **¿Se ha puesto en marcha alguna de las siguientes medidas, algunas de las cuales son obligatorias según el Reglamento General de Protección de Datos (RGPD) (Samper, 2020), o un equivalente no europeo?**
 - **Evaluación del impacto de la protección de datos (DPIA)**

No, debido a que nuestra recopilación de datos fue realizada sobre únicamente 12 personas las cuales dieron su consentimiento.

- No llevamos a cabo una evaluación sistemática y exhaustiva de aspectos personales de personas físicas basada en un tratamiento automatizado, como elaboración de perfiles sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas.
- No llevamos a cabo un tratamiento a gran escala de datos personales que revelen origen étnico o racial, opiniones políticas, convicciones religiosas o filosóficas, afiliación sindical, datos genéticos, biométricos, datos relativos a la salud, vida u orientación sexuales, o datos personales relativos a condenas e infracciones penales.
- No realizamos una observación sistemática a gran escala de una zona de acceso público.

- No tratamos datos de menores.
- No pone en riesgo a las personas si sus datos son filtrados.

- **Designación de un responsable de la protección de datos (DPO) e inclusión de este en una fase temprana del desarrollo.**

Sí, actualmente todos actuamos bajo el rol de DPO.

- **Mecanismos de supervisión del tratamiento de los datos (incluida la limitación del acceso al personal cualificado, los mecanismos de registro del acceso a los datos y la realización de modificaciones).**

Sí, todos los datos están almacenados de manera que solo pueden ser accesibles, tratados y modificados por los miembros de la investigación, como se comunicó a cada uno de los voluntarios previamente a la realización del experimento. La modificación y acceso a los datos queda registrada.

- **Medidas para lograr la privacidad por diseño y por defecto (por ejemplo, cifrado, seudonimización, agregación, anonimización).**

Sí, los datos recogidos para su tratamiento son anónimos y solo visibles para el docente que haga uso de nuestra herramienta. Además, todos los datos son eliminados tras su tratamiento.

- **Minimización de datos, en particular de los datos personales (incluidas las categorías especiales de datos).**

Sí, al haber creado una herramienta con el objetivo de ser lo menos intrusiva posible, solo necesitamos la cara de la persona para que nuestra herramienta funcione.

- **¿Se ha incorporado el derecho a retirar el consentimiento, el derecho de objeción y el derecho a ser olvidado en el desarrollo del sistema de IA?**

Sí, previamente a la recogida de datos, pusimos en conocimiento de cada uno de los participantes del experimento, su derecho a retirar el consentimiento en cualquier momento, objetar, y ser olvidado. De manera que eliminaríamos todos sus datos de ser solicitado por la persona.

- **¿Se ha tenido en cuenta las implicaciones para la privacidad y la protección de datos de los datos recogidos, generados o procesados durante el ciclo de vida del sistema de IA?**

Si, desde un primer momento pusimos en conocimiento de los voluntarios que participaron en el experimento sus derechos sobre estos datos y el tratamiento, almacenamiento y protección que íbamos a llevar a cabo. Debido a esto, hemos realizado la anonimización de los datos que trata la IA, restringido el acceso de los datos almacenados a solamente personal autorizado y eliminado los nuevos datos generados o procesados con el objetivo de garantizar la privacidad de la persona.

- **¿Se ha tenido en cuenta las implicaciones para la privacidad y la protección de datos de los datos de entrenamiento no personales del sistema de IA o de otros datos no personales procesados?**

Si, toda información almacenada sobre la puntuación de los vídeos y la decisión de continuar con la visualización, están almacenados de manera restringida y anonimizada, solamente accesible por los miembros de la investigación.

- **¿Se ha adaptado el sistema de IA a las normas pertinentes (por ejemplo, ISO (2017), IEEE (2021)) o a los protocolos ampliamente adoptados para la gestión y el gobierno de los datos (diarios)?**

No, al ser el desarrollo de un pequeño prototipo no se han seguido los protocolos establecidos. Aun así, se ha optado siempre por la protección de la privacidad y los datos tratados, además de solo utilizarse en situaciones controladas.

9.4.4. Transparencia

Trazabilidad

- **¿Se han establecido medidas que aborden la trazabilidad del sistema de IA durante todo su ciclo de vida?**

No aplica, al ser un prototipo no se ha establecido un ciclo de vida del software.

- **¿Se han establecido métricas para evaluar continuamente la calidad de los datos introducidos en el sistema de IA?**

Sí, los datos recogidos han sido evaluados frame por frame, recortado caras y eliminando los frames anómalos, se explica con detalle en el apartado 5.

- **¿Se puede rastrear los datos utilizados por el sistema de IA para tomar una decisión o recomendación determinada?**

No aplica ya que no toma decisiones ni recomienda, simplemente informa.

- **¿Se puede rastrear qué modelo o reglas de IA condujeron a las decisiones o recomendaciones del sistema de IA?**

No, al ser un prototipo no se han trazado las decisiones de la IA (Los patrones de aburrimiento e interés no se ven explícitamente en las capas de la R. Neuronal)

- **¿Se han establecido métricas para evaluar continuamente la calidad de los resultados del sistema de IA?**

No aplica, al ser un prototipo que tiene un 40 % de precisión aproximadamente, es lógico que la calidad de los resultados no sea la óptima. En un futuro las métricas de evaluación deberían separar los resultados por género y razas, sólo para informar a los desarrolladores si existe una tasa de error anormalmente alta y sesgada.

- **¿Se han establecido prácticas adecuadas para registrar las decisiones o recomendaciones del sistema de IA?**

No aplica, es un prototipo que tiene un 40 % de precisión aproximadamente.

Explicabilidad

- **¿Se han explicado las decisiones del sistema de IA a los usuarios?**

No, aunque sería interesante añadir una funcionalidad que permita mostrar por qué se ha llegado a una conclusión u otra, por ejemplo, si el 80 % de la clase está aburrida, mostrar lo que ha conformado este porcentaje es un 40 % del alumnado con los ojos cerrados, un 30 % con la cabeza apoyada y un 10 % distraído.

- **¿Se consulta continuamente a los usuarios si entienden las decisión del sistema de IA?**

No por lo anteriormente mencionado.

Comunicación

- **En los casos de sistemas interactivos de IA (por ejemplo, chatbots, robo-abogados), ¿se comunica a los usuarios que están interactuando con un sistema de IA en lugar de con un humano?**

No aplica.

- **¿Se han establecido mecanismos para informar a los usuarios sobre la finalidad, los criterios y las limitaciones de las decisiones generadas por el sistema de IA?**

No aplica, esta herramienta es un prototipo, la finalidad es informar sobre el aburrimiento y las limitaciones son un 40 % de precisión, por lo cual es muy poco fiable. Además los usuarios que la han testado siempre han estado bajo la supervisión de un desarrollador del proyecto, por lo cual, en el momento de tener alguna duda, se ha explicado verbalmente.

- **¿Se ha comunicado a los usuarios las ventajas del sistema de IA?**

Sí, actualmente todos los usuarios que han probado el prototipo han sido informados de las ventajas futuras de la herramienta.

- **¿Se ha comunicado a los usuarios las limitaciones técnicas y los riesgos potenciales del sistema de IA, como su nivel de precisión y/o sus tasas de error?**

Sí, actualmente todos los usuarios que han probado el prototipo han sido informados de las limitaciones y riesgos de la herramienta, como su 40 % de precisión.

- **¿Se ha proporcionado a los usuarios material de formación adecuado y advertencias sobre cómo utilizar adecuadamente el sistema de IA?**

Sí, la herramienta ha sido testada por los usuarios bajo la supervisión de los desarrolladores, se conocía el funcionamiento y se guiaba al usuario.

9.4.5. Diversidad, equidad y no discriminación

Evitar sesgos injustos

- **¿Se ha establecido una estrategia o un conjunto de procedimientos para evitar crear o reforzar un sesgo injusto en el sistema de IA, tanto en lo que respecta al uso de los datos de entrada como al diseño del algoritmo?**

En nuestro caso, el sesgo injusto se crearía por el dataset que hemos hecho. La única estrategia que hemos seguido ha sido la de crear dicho dataset con un 50 % de hombres y un 50 % de mujeres. Es probable que la precisión disminuya con diversas razas, esto se debe a que no hemos podido recolectar más información en el dataset, aunque somos conscientes de ello.

- **¿Se ha tenido en cuenta la diversidad y la representatividad de los usuarios finales de los datos?**

Se ha tenido en cuenta pero no se ha podido recolectar información más diversa.

- **¿Se ha probado para grupos específicos o casos de uso problemáticos?**

No se ha probado por lo anteriormente mencionado.

- **¿Se ha investigado y usado herramientas técnicas disponibles públicamente, que sean de última generación, para mejorar su comprensión de los datos, el modelo y el rendimiento?**

No, toda la fase de investigación ha estado enfocada en estudiar las emociones, el aburrimiento, el interés y su posible detección.

- **¿Se ha evaluado y puesto en marcha procesos para probar y controlar los posibles sesgos durante todo el ciclo de vida del sistema de IA (por ejemplo, sesgos debidos a posibles limitaciones derivadas de la composición de los conjuntos de datos utilizados (falta de diversidad, falta de representatividad))?**

No aplica, al ser un prototipo el ciclo de vida es prácticamente nulo.

- **¿Se ha tenido en cuenta la diversidad y la representatividad de los usuarios finales de los datos?**

Se ha tenido en cuenta, y hemos intentado que sea lo más diverso dentro de nuestras posibilidades.

- **¿Se han puesto en marcha iniciativas de educación y concienciación para ayudar a los diseñadores y desarrolladores de IA a ser más conscientes del posible sesgo que pueden inyectar en el diseño y desarrollo del sistema de IA?**

No aplica al no ser una empresa y al no tener capital para poner en marcha dichas iniciativas.

- **¿Se ha garantizado un mecanismo que permita señalar los problemas relacionados con la parcialidad, la discriminación o el mal funcionamiento del sistema de IA?**

Sí, actualmente todos los usuarios han sido informados verbalmente de dichos problemas y se han resuelto dudas, en el caso de un proyecto futuro, sería interesante tener una web en la que se resuelvan problemas y dudas comunes, además de un formulario de contacto para resolver otro tipo de problemas.

- **¿Se han establecido pasos y formas de comunicación claros sobre cómo y a quién se pueden plantear estas cuestiones?**

Sí, al ser un prototipo se ha informado verbalmente a los usuarios.

- **¿Se ha identificado a los sujetos que podrían verse afectados directa o indirectamente por el sistema de IA, además de los usuarios finales?**

Sí, se han identificado. Los afectados directamente son el alumnado que está siendo grabado y el profesor que está dando la clase. Los afectados indirectamente son los familiares o conocidos del alumnado, que pueden ser grabados por la cámara.

- **¿La definición de equidad se utiliza habitualmente y se aplica en cualquier fase del proceso de creación del sistema de IA?**

Sí, en todo momento hemos tenido clara la necesidad de tener una inteligencia artificial equitativa y nos hemos enfocado en conseguirlo de la mejor manera posible.

- **¿Se han considerado otras definiciones de equidad antes de elegir ésta?**

La equidad es que nadie salga favorecido si ello implica perjudicar a un tercero, en este caso, nuestra definición de equidad es que todos los usuarios puedan tener un porcentaje muy similar de precisión en nuestra herramienta.

- **¿Se ha consultado a las comunidades afectadas sobre la definición correcta de equidad, por ejemplo, a los representantes de las personas mayores o de las personas con discapacidad?**

No, actualmente no se ha consultado a las comunidades afectadas.

- **¿Se ha garantizado un análisis cuantitativo o una métrica para medir y probar la definición de equidad aplicada?**

No existe una métrica que nos permita medir la equidad actual de nuestro proyecto

- **¿Se han establecido mecanismos para garantizar la equidad en su sistema de IA?**

No se han establecido mecanismos porque tampoco se han establecido métricas y no podríamos medir el impacto de dichos mecanismos.

Accesibilidad y diseño universal

- **¿Se ha asegurado que el sistema de IA se ajusta a la variedad de preferencias y capacidades de la sociedad?**

No, actualmente la herramienta es un prototipo por lo que no es tan específica.

- **¿Se ha evaluado si la interfaz de usuario del sistema de IA es utilizable por las personas con necesidades especiales o discapacidades o con riesgo de exclusión?**

No, actualmente la herramienta es un prototipo por lo que no es tan específica.

- **¿Se ha asegurado de que la información y la interfaz de usuario del sistema de IA sean accesibles y utilizables también para los usuarios de tecnologías de asistencia (como los lectores de pantalla)?**

No, actualmente la herramienta es un prototipo por lo que solo pueden usarlo los desarrolladores de la herramienta en sus ordenadores.

- **¿Se ha involucrado o consultado a los usuarios finales que necesitan tecnología de apoyo durante la fase de planificación y desarrollo del sistema de IA?**

Los usuarios finales de la herramienta son el docente y el alumnado. Nosotros al ser estudiantes hemos considerado que la herramienta sería útil, y Borja, al ser director del proyecto y docente, también considera que su uso podría ser interesante, por lo que sí, ambos tipos de usuario han estado involucrados en ambas fases.

- **¿Se ha asegurado que los principios del diseño universal se tuvieran en cuenta en cada paso del proceso de planificación y desarrollo, si fuera el caso?**

Al no ser una herramienta invasiva, es decir, carecer de sensores específicos y físicos conectados al cuerpo, y depender de una cámara, consideramos que tiene un diseño parcialmente universal, actualmente no está enfocado en un diseño accesible para personas con discapacidades y tampoco para personas que carezcan de cámaras.

- **¿Se ha tenido en cuenta el impacto del sistema de IA en los posibles usuarios finales?**

Sí, existe la posibilidad de que un órgano educativo pueda implementar una solución basada en esta herramienta, y existirán docentes e incluso alumnos/tutores de los alumnos que no estén de acuerdo con dicho uso, por lo que su uso debería ser completamente opcional para los usuarios finales.

- **¿Se ha evaluado si el equipo que ha participado en la construcción del sistema de IA se ha comprometido con los posibles usuarios finales?**

Sí, lógicamente estamos comprometidos con la herramienta porque al igual que somos sus desarrolladores, también somos estudiantes.

- **¿Se ha evaluado si hay grupos que podrían verse afectados de forma desproporcionada por los resultados del sistema de IA?**

No se ha evaluado, pero es posible que personas de distintas razas puedan verse afectadas ya que el dataset de entrenamiento no ha tenido los suficientes participantes, ni tampoco ha tenido diversidad de grupos.

- **¿Se ha evaluado el riesgo de que el sistema sea injusto para los colectivos de usuarios finales?**

Actualmente el prototipo no ha evaluado dicho riesgo por lo anteriormente comentado.

Participación de las partes interesadas

- **¿Se ha considerado un mecanismo para incluir la participación del mayor número posible de interesados en el diseño y desarrollo del sistema de IA?**

No porque la herramienta es un prototipo, aunque, en caso de seguir con el proyecto, queremos que el dataset tenga una amplia mayoría de diversos estudiantes, además de la supervisión de varios docentes e investigadores que permitan pivotar la herramienta.

9.4.6. Bienestar social y ambiental

Bienestar ambiental

- **¿Existen posibles impactos negativos del sistema de IA en el medio ambiente?**

No, nuestra IA es un prototipo pequeño y no usa muchos recursos.

- **¿Qué impactos potenciales se identifican?**

No hemos detectado que tenga un impacto potencial.

- **En la medida de lo posible, ¿Se han establecido mecanismos para evaluar el impacto medioambiental del desarrollo, el despliegue y/o el uso del sistema de IA (por ejemplo, la cantidad de energía utilizada y las emisiones de carbono)?**

Aunque no se haya establecido un mecanismo directo para medir el posible impacto medioambiental, si que la IA está diseñada de la forma más eficiente para consumir el menor número de recursos posibles

- **¿Se han definido medidas para reducir el impacto medioambiental del sistema de IA a lo largo de su ciclo de vida?**

No se ha contemplado porque no se ha visto necesario.

Impacto en el trabajo y las habilidades

- **¿Influye el sistema de IA en el trabajo humano y en los acuerdos laborales?**

No, ya que la IA es una herramienta externa para la ayuda del docente en la enseñanza.

- **¿Se ha preparado el camino para la introducción del sistema de IA en su organización informando y consultando previamente a los trabajadores afectados y a sus representantes (sindicatos, comités de empresa (europeos))?**

Sí, hemos preguntado a diversos profesores sobre los posibles impactos de la herramienta y preparado un camino para el uso de esta.

- **¿Se han adoptado medidas para garantizar que se entiendan bien las repercusiones del sistema de IA en el trabajo humano?**

Sí, los usuarios están informados de que la herramienta solo sirve de apoyo al docente.

- **¿Se ha asegurado de que los trabajadores entienden cómo funciona el sistema de IA, qué capacidades tiene y cuáles no?**

Sí, en esta misma memoria se detalla con exactitud las capacidades de la herramienta.

- **¿Podría el sistema de IA crear el riesgo de descualificación de la mano de obra?**

No, ya que es una herramienta diseñada para el apoyo del docente, no para sustituirlo.

- **¿Se han tomado medidas para contrarrestar los riesgos de descualificación?**

Sí, la herramienta fue planteada y diseñada para un uso exclusivo de apoyo al docente, no para sustituir ninguna de sus funciones.

- **¿Se promueve o requiere el sistema nuevas competencias (digitales)?**

No es necesario, la herramienta está diseñada de una forma de fácil uso.

- **¿Se han proporcionado oportunidades y materiales de formación para la recualificación y el perfeccionamiento?**

No aplica, ya que no ha sido necesario aprender nuevas competencias.

Impacto en la sociedad y la democracia

- **¿Podría el sistema de IA tener un impacto negativo en la sociedad en general o en la democracia?**

No tendrá un impacto dado que no se va a usar fuera de un grupo muy reducido de personas.

- **¿Se ha evaluado el impacto social del uso del sistema de IA más allá del (usuario final) y del sujeto, como las partes interesadas potencialmente afectadas de forma indirecta o la sociedad en general?**

Sí, pero dado al número reducido de usuarios, no habrá efectos indirectos.

- **¿Se han tomado medidas para minimizar los posibles daños sociales del sistema de IA?**

Si, la herramienta está diseñada para ser lo menos intrusiva y que sea exclusivamente usada para el apoyo del docente.

- **¿Se han tomado medidas para garantizar que el sistema de IA no tenga un impacto negativo en la democracia?**

No tendrá un impacto dado que no se va a usar fuera de un grupo muy reducido de personas.

9.4.7. Responsabilidad

Auditabilidad

- **¿Se han establecido mecanismos que faciliten la auditabilidad del sistema de IA (por ejemplo, la trazabilidad del proceso de desarrollo, la obtención de datos de entrenamiento y el registro de los procesos, resultados e impactos positivos y negativos del sistema de IA)?**

Sí, esta misma memoria y el paper del anexo detalla todo el proceso de desarrollo y de la obtención del dataset.

- **¿Se ha asegurado de que el sistema de IA pueda ser auditado por terceros independientes?**

Si, ya que tanto esta memoria como el paper y el código está disponible para todo el mundo

Gestión de riesgos

- **¿Se ha previsto algún tipo de orientación externa o procesos de auditoría de terceros para supervisar las cuestiones éticas y las medidas de responsabilidad?**

No, no se ha planificado ninguna orientación externa dada la escala del proyecto.

- **¿La participación de estos terceros va más allá de la fase de desarrollo?**

No aplica dado la nula participación de terceros.

- **¿Se organizó una formación sobre los riesgos y, en caso afirmativo, informa también sobre el posible marco jurídico aplicable al sistema de IA?**

No se organizó ninguna formación dado que es un prototipo por lo que genera pocos o ningún riesgo.

- **¿Se ha considerado la posibilidad de crear una junta de revisión ética de la IA o un mecanismo similar para debatir las prácticas generales de responsabilidad y ética, incluidas las posibles zonas grises poco claras?**

No se consideró dado que es un prototipo de una escala pequeña.

- **¿Se ha establecido un proceso para debatir y supervisar y evaluar de forma continua la adhesión del sistema de IA a esta Lista de Evaluación**

para una IA digna de confianza (ALTAI)? (Hochrangige Expertengruppe für Künstliche Intelligenz (HEG-KI), 2020)

No se ha establecido ningún debate dado que no se va a usar fuera de un grupo muy reducido de personas

- **¿Se incluye en este proceso la identificación y documentación de los conflictos entre los 6 requisitos mencionados o entre diferentes principios éticos y la explicación de las decisiones de compensación"tomadas?**

No aplica dado que no ha habido el proceso anterior

- **¿Se ha proporcionado una formación adecuada a las personas que participan en dicho proceso y abarca también el marco jurídico aplicable al sistema de IA?**

No aplica dado que no ha habido el proceso anterior

- **¿Se ha establecido un proceso para que terceras partes (por ejemplo, proveedores, usuarios finales, sujetos, distribuidores/vendedores o trabajadores) informen de posibles vulnerabilidades, riesgos o sesgos en el sistema de IA?**

No se ha establecido dado que es un prototipo el cual el grupo de usuarios va a ser muy reducido.

- **¿Se fomenta este proceso la revisión del proceso de gestión de riesgos?**

No aplica la revisión del proceso

- **Para las aplicaciones que pueden afectar negativamente a las personas, ¿se han establecido mecanismos de reparación por diseño?**

No aplica porque los usuarios van a ser un grupo muy reducido y tienen contacto directo con nosotros.

Anexo 2 - Vital Emo

Vital Emo: The boredom detector with a machine learning perspective.

Diego ISAR MUÑOZ^a, Álvaro ORTIZ MARCHUT^a, Daniel PADILLA RODRIGUEZ^a, Sofía PRIETO IBÁÑEZ^a, Meriem EL YAMRI^a, Josefa ROS VELASCO^a, Borja MANERO^a

^a*Complutense University of Madrid, Spain*

*diegisar@ucm.es, alvort03@ucm.es, danpad01@ucm.es, sofiapri@ucm.es, melyamri@ucm.es, josros@ucm.es, bmanero@ucm.es

Abstract: Boredom is one of the main problems found in 21st-century educational environments. Today's society is hyperconnected and exposes us to a large number of daily stimuli. This daily overstimulation can negatively affect students who are facing several hours of online classes (especially in the new context derived from COVID-19). In these classes, students experience a shortage of stimuli, resulting in boredom and a shorter attention span. Past studies have shown that certain facial patterns help detect lack of engagement and boredom. The most noticeable are the widening of the eyelids, the opening of the lips, and the gaze direction. As for posture, there are references to using the hand to support the head, maintaining a reclining posture, or hiding part of the face. Its counterpart, interest, can be reflected in patterns such as maintaining a wide-eyed gaze, smiling, or maintaining an upright posture.

This paper presents the development, architecture, and evaluation of a tool called VITAL EMO. Through artificial intelligence (trained with facial patterns), this tool seeks: 1) to detect, through a camera, students' boredom, and 2) to alert teachers of this situation, allowing them to change the methodology or strategy to reengage the students on the spot.

The creation of the dataset used to train the AI was carried out with 12 people. The participants took part telematically; each were recorded as they watched a video and answered questions about it.

Despite using a small dataset, the study results show that it is possible to detect facial patterns associated with engagement and boredom, such as smiling, squinting, facial occlusion, or avoiding being directly in front of the camera. These results pave the way for future research with a larger dataset to increase the effectiveness of AI and further develop the VITAL EMO tool.

Keywords: Boredom, emotions, education, students, classes, camera, video, detector, artificial intelligence, machine learning, deep learning, convolutional networks, computer application, vital emo.

1. Introduction

Living in a hyperconnected society results in a need for constant stimulation. This is clearly evidenced in the educational environments of developed countries. Despite the efforts made by researchers and teachers to integrate technology into the classroom and to update teaching methodologies to make teaching more attractive, such endeavors do not always result in greater student engagement. This is partly due to the fact that educational environments, to date, have not been able to keep up with the pace of over-stimulation and over-excitement to which students are exposed in their daily lives outside the educational institution. The immediate aftermath of this imbalance is the emergence of boredom and short attention spans.

According to the Meaning and Attentional Components Model (MAC) developed by Erin C. Westgate and Timothy D. Wilson, boredom is a state of displeasure resulting from "(a) an attentional component, namely mismatches between cognitive demands and available mental resources, and (b) a meaning component, namely mismatches between activities and valued goals (or the absence of valued goals altogether)" (2018, p. 689). In other words, boredom is an unpleasant state that "is experienced when people feel either unable or unwilling to cognitively engage with their current activity" (2018, p. 693; see also Eastwood et al. 2012).

Several studies have addressed the issue of boredom in educational settings (Sharp et al., 2017, 2019; Goetz and Hall, 2014; Mann, 2009; Xie et al. 2021; Tze et al., 2016; Nett et al., 2011; Finkielstein, 2020). However, almost none have focused on boredom in online teaching, despite the fact that this teaching mode is the most widespread worldwide today due to the prevention measures implemented in the face of the COVID-19 pandemic. In this regard, there is a lack of resources in research and teaching applications to analyze the value of emotions, and specifically boredom, in the area of online education, as pointed out by psychologist Reinhard Pekrun (2005). Among the few approaches to this issue is the research conducted by Rebollo-Catalán et al. (2008). In their paper, the authors sought to better understand student circumstances in a blended learning environment. Rebollo-Catalán and her team measured the students' feelings in order to learn and evaluate the emotions associated with online learning, including boredom.

If, as stated by Rebollo-Catalán et al. (2008), boredom is one of the emotions involved in online learning, and this leads to the learner's disengagement with the subject due to hypostimulation resulting from an environment that is unable to maintain optimal levels of arousal, the question that both experts in boredom studies and pedagogues must seek to answer is how to objectively and reliably detect that students are getting bored in online classes and how to share this information with the teacher so that he or she can redirect the lesson to re-engage students on the spot and thus prevent them from disconnecting or seeking stimulation in other common places (social networks, video games, web pages, smartphones...) during class.

In order to find out if students are getting bored in online classes and to provide the teacher the means to react to this, it is necessary to find a formula that allows us to objectify and quantify boredom. Experts such as the Canadian researcher James Danckert and his team at The Danckert's Lab have been working for years on measuring boredom through objective parameters, that is, beyond the classic scales of subjective measurement of the experience of boredom (BPS - Boredom Proneness Scale [Farmer and Sundberg, 1986]). Many of Danckert and colleagues' experiments have focused on employing fMRI techniques for capturing neuroimaging of the brain to measure boredom (Isacescu, Struck, & Danckert, 2017; Danckert & Merrifield, 2018; Danckert, 2019). A paper published by Danckert and Merrifield (2014) explored the most obvious physiological correlate of boredom by employing measurements of heart rate, electrodermal activity, and cortisol levels. The outcomes were variable depending on the person and their tendency to boredom.

These methods for measuring boredom turned out to be quite intrusive and were not intended to be applied in the aforementioned context, so it seems appropriate to find a way to immediately detect boredom without conditioning students. In this regard, other studies have shown that there are facial patterns that help detect whether a person is engaged or bored. Silvia (2001) noted that the most perceptible patterns showing interest were head stillness, a widening of the eyelids, a slight parting of the lips, and a more fixed gaze. As for body posture, Kroes (2007) associated boredom with resting the chin in one's hand, a reclined sitting posture, or hiding part of the face. Its counterpart, interest, can be seen in some patterns such as keeping a wide-eyed gaze, smiling, or maintaining an upright posture.

In relation to the above, in this paper we present the development, architecture, and evaluation of the tool VITAL EMO, whose main goal is to detect boredom in online educational environments. In the following sections, we will present (1) the architecture followed for the creation of this tool based on AI training for the detection of facial patterns related to boredom, (2) the methodology used for the training and evaluation of the tool, (3) the results of the evaluation, and (4) a discussion about the limitations of the present study and the approach to future work.

2. Proposed Architecture

To address the described problem from a technological perspective and being as less intrusive as possible with the user, we decided to effect the detection of boredom through facial recognition (Schroff et al., 2015); that is to say, through a camera that captures the user's facial expression and applying artificial intelligence to detect patterns (Zhang et al., 2019) related to boredom.

The selected AI model is a convolutional neural network that classifies an image into classes through the repetition of a pattern.

For example, a convolutional neural network (O'Shea & Nash, 2015) is able to detect a face through patterns such as the eyes or the mouth. In this case, we wanted it to be able to detect facial expressions related to boredom or engagement.

Specifically, we have proposed a model based on VGG16 (Simonyan & Zisserman, 2015) where the input images to the neural network have a dimension of 48x48 pixels in grayscale. The following pattern of layers is applied to these images three times:

Convolution > Batch Norm. > Convolution > Batch Norm. > Max pool. > Dropout

- Convolution (C): Multiplication of a part of the image by a matrix called filter.
- Batch Normalization (BN): Rescaling the data, speeding up the neural network.
- Max pooling (MP): Reduction of the matrix by dividing it into equal regions, where the maximum value of each region is chosen.
- Dropout (DR): Layer that ignores randomly selected neurons.
- Fully Connected (F): Connection of all neurons of a layer with its successor.

After the application of this pattern, the information passes through two Fully Connected layers. The output layer generates an output of 7 classes: anger, disgust, fear, happiness, sadness, surprise and neutral.

After training, we apply Transfer Learning. Therefore, we reset the values of the last 10 layers and, in the last one, we set an output of three classes (bored, neutral, engaged) instead of the previous seven output classes.

Finally, we retrained the last 10 layers of the neural network (Figure 1).

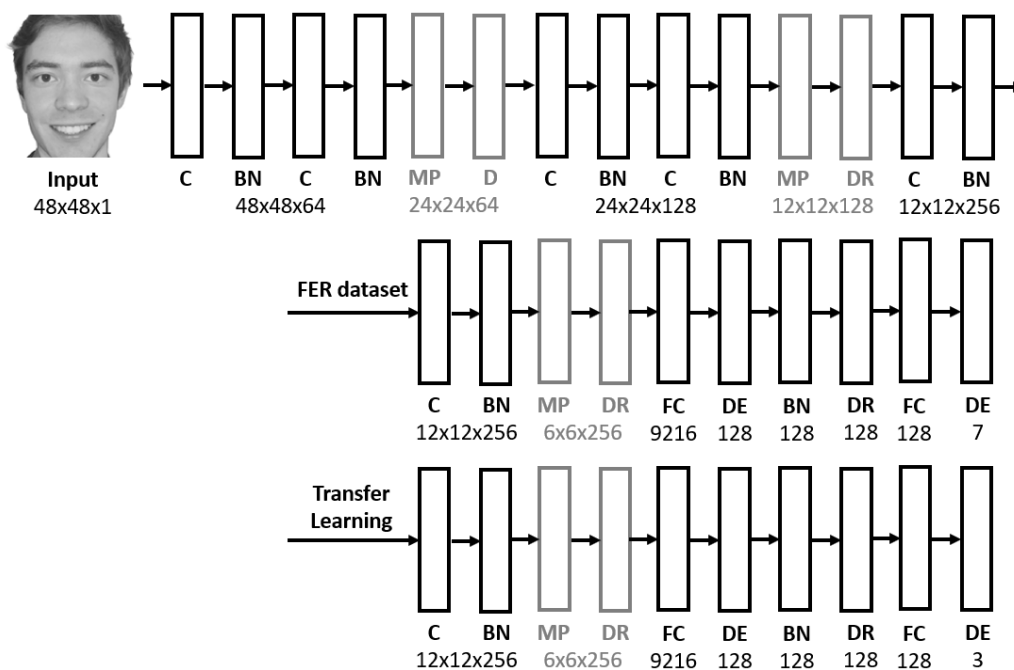


Figure 1. Project's architecture.

Neural networks are composed of nodes that simulate the behavior of the human brain. These nodes are connected to each other by values called weights. The term 'training' refers to the variation of these weights, in order to reduce the error of transforming an input into an output.

A specific case is convolutional neural networks, which through smaller and distinct matrices called filters, allows to find patterns in an image.

The images we have used to train the artificial intelligence have been extracted from two datasets: Kaggle's dataset and our own dataset.

Facial Expression Recognition (FER) Challenge from Kaggle (FER Challenge, n.d.) is a dataset that has helped us recognize faces and emotions from images and their respective labels. This dataset has approximately 36,000 rows, and is divided into the following columns:

- *emotion*: integer that represents the emotion expressed in the image, with 0 = Angry, 1 = Disgust, 2 = Fear, 3 = Happy, 4 = Sad, 5 = Surprise, and 6 = Neutral.
- *usage*: Training, Public Test and Private Test.
- *pixels*: 2304 integers from 0 to 255. Represents a pixel in grayscale.

We have extended this dataset to have more training data applying Data Augmentation (*Keras API*, n.d.) techniques (i.e., rotation, shifting, zooming) using the Keras API.

For training, the dataset has been divided into 90% train set and 10% test set. These data serve as the basis for the convolutional network to be able to detect faces and small emotion features.

In our research, we have not found a public dataset that specifically differentiates between boredom, neutral state and engagement. For this reason, we decided to create our own dataset. In the 'Methodology' section we explain the procedure that we followed to obtain the videos for the dataset.

Each of these videos has been post-processed:

1. Division of the video into frames.
2. Face detection and cropping
3. Conversion to grayscale
4. Size reduction to 48x48x1 using OpenCV1 (*OpenCV Docs*, n.d.).
5. CSV file generation with the user identifier, the video identifier, the frame identifier, the pixels and the engagement score felt at the moment of watching the video (provided by the user).

The dataset has been divided by the person's identifier. This has been done because mixing all the frames could provide false hit rates in training and cross-validation that do not correspond to the hit values in a real environment. The split of this dataset has been 70% of individuals to training and 30% of individuals to cross-validation.

We have retrained the artificial intelligence with this dataset through Transfer Learning. We have retrained the artificial intelligence with this dataset through Transfer Learning.

3. Methodology

3.1 Participants

The experiment involved a total of 12 participants (50% men and 50% women) between the ages of 18 and 56 years old and of different nationalities. 5 participants were students, while the rest were professionals. Education-wise, the participants' education ranged from compulsory education to university. Most of the participants had Spanish as their native language and enough knowledge of English to understand a conversation or a monologue.

3.2 Experimental design

All participant recordings were conducted telematically. Each participant was notified that the recording was about to begin, and they had a briefing on the purpose and steps of the experiment. Once informed, they signed a legal agreement of consent to participate in the experiment. Afterwards, we requested some personal data such as name, level of studies or native language.

Subsequently, each participant watched a video and filled in a questionnaire about it. This last process was repeated for each video until the last one or until the participant wanted to stop or take a break (Figure 2).

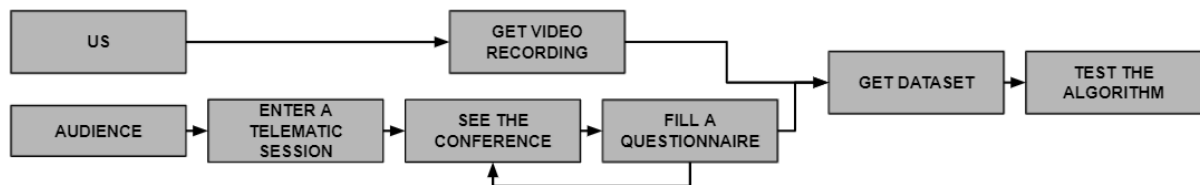


Figure 2. Methodology.

3.3 Instruments

The video set shown to the participants contains 19 videos (duration: 48'' - 3') extracted from YouTube with Creative Commons license. All the videos, except two, are fragments of lectures or classes. The videos were shown in a random order to each participant.

We used different online tools to record the participants: Google Meet; Discord and Microsoft Teams, among others, selecting the most appropriate one depending on the availability and resources of each participant.

We used two questionnaires created with Google Forms. The first questionnaire was a pre-experiment and it was intended to collect personal data, so that we could obtain statistics of the participants based on the information collected. The data collected is shown in Table 1.

Table 1. *Personal information*

Measured variable	Type	Range
Name	Text	-
Gender	Option	Male, Female, Other
Age	Number	18 - 99
Nationality	Text	-
Current job	Text	-
Languages	Text	-
Highest level of completed education	One choice	Elementary school, middle school, high school, higher education, university, Master's degree, doctoral degree...

The second questionnaire was done post-experiment and it was intended to collect data on the state of boredom after watching each one of the videos.

The questions in the latter questionnaire consisted of a score ranging from 1 (very boring) to 9 (very interesting) to be assigned by the participant to each video, and a question indicating whether the person would be interested in continuing to watch the video after having seen the fragment shown (see Table 2).

Table 2. *Video punctuation*

Measured variable	Type	Range
Boredom	One choice	1 (boring) - 9 (interesting)
Continue watching	Yes or No	0 - 1

4. Results

After recording all the participants during the experiment and processing the videos, we obtained approximately 64,000 frames from 12 different participants with their corresponding score.

We obtained 226 total responses from the post-experiment questionnaire, the values of which were distributed among the different options of the questionnaire, as seen in Figure 3.

As we anticipated, there was a greater number of scores with values 1 (boring) and values 7 and 8 (values very close to engaged), verifying that the selection of videos we made induced emotions closer to the edge, rather than neutral ones.

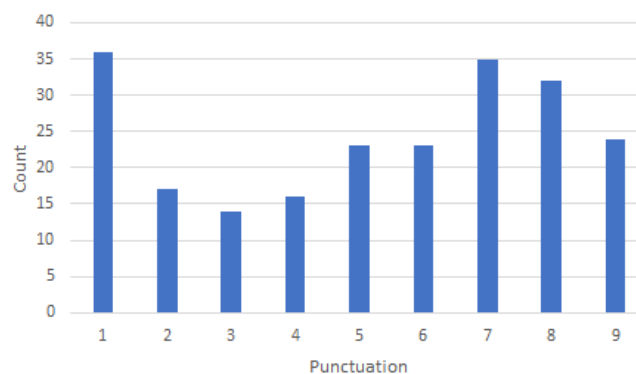


Figure 3. Video score results.

The results we obtained from the second question of the post-experiment questionnaire showed an almost equal distribution between the videos rated positively and negatively (56.2% answered that they wanted to continue watching the video, compared to 43.8% who did not).

The training with the Kaggle dataset generated an emotion prediction accuracy rate of around 70%.

After the first training, we applied transfer learning to detect facial features characteristic of emotions that serve to predict boredom.

The graphs in Figure 4 shows the progress of the training process with the second dataset, where we can observe how the precision with the training data increases while the validation data remains stable between 30% and 40% of accuracy.

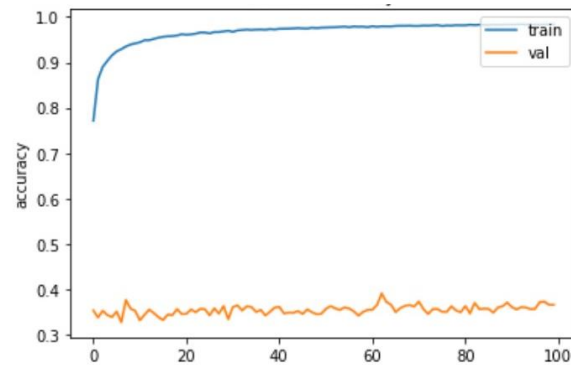


Figure 4. Model accuracy.

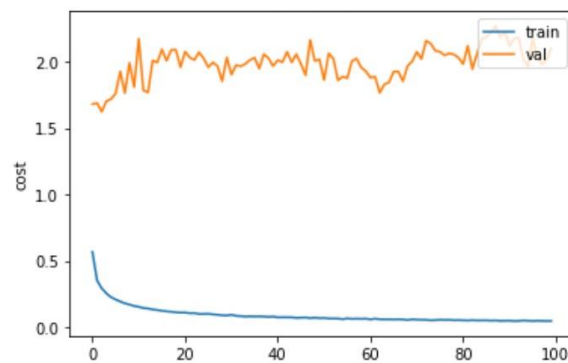


Figure 5. Model cost.

Based on these results, we can observe an overfitting effect of the neural network, because despite the fact that the input consists of a large set of tagged frames, most of them are very similar to each other (they belong to the same participant).

However, the classifier managed to predict certain patterns that are symptomatic of boredom or interest. On the one hand, during the testing process, we found that the application detected boredom traits when the participants rested their head on their hand, squinted their eyes or turned their head away from the camera that was recording. On the other hand, we noticed how other gestures such as smiling, keeping the eyes wide open or facing the camera directly were perceived as engagement by the classifier.

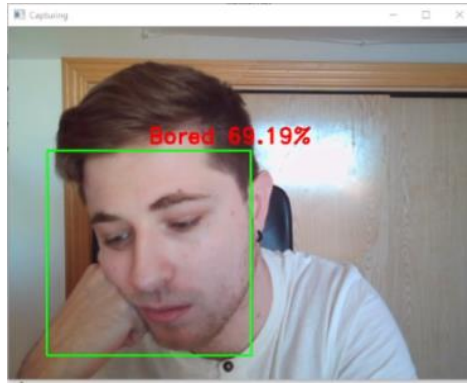


Figure 6. Example of boredom recognition.

After multiple training sessions, the patterns detected varied depending on the dataset part selected for training and validation. Even then, some patterns such as holding the head or smiling, were consistent over several training sessions.

5. Limitations

The generated dataset is not large nor varied enough, due to time constraints and limited resources available. We were, however, aware from the beginning of the investigation that this could cause an overfitting problem in the artificial intelligence.

Another limitation worth mentioning is that the hardware material available for this work was not powerful enough to quickly iterate over the volume of data and to support the architecture of the project.

6. Conclusions and future work

The objective of this paper is to present the architecture and the evaluation of Vital Emo, a tool that offers teachers real-time information related to the experience of boredom of their online students, so that they can redirect their teaching strategy on the spot.

Online students' faces are sufficient to detect characteristic patterns of boredom or engagement but not to determine their degree with accuracy. Our artificial intelligence has managed to classify these states from explicit gestures, i.e., facial occlusion caused by the hand, smiling, having the eyes wide open, or facing the camera directly. According to the literature, these gestures represent typical boredom and engagement signals.

In the case of holding the head with the hand, the trained AI tends to recognize boredom more easily when the person uses the left hand. We understand that this is due to a reduced dataset in which we have few images of people using the right hand to hold their head even though both are clear patterns of boredom.

It is also worth noting the importance of lighting in detecting face expressions. In some tests with team members, we observed how those with worse lighting obtained a higher number of false positives.

In the future, we will increase the training and testing dataset and the types of collected data (for instance, including body language). Furthermore, we will consider migrating the project to TensorFlow since it allows a more detailed modification of the parameters and adds a precise recurrence to the neural network. Finally, it would also be advisable to add functionalities to the tool in a way that would allow information about online students to reach the teacher, such as adding the tool to some educational platforms like Collaborate or Zoom, since it currently works locally with only one face.

Acknowledgements

This project has been funded by the Ministry of Science, Innovation and Universities of Spain (Didascalías, RTI2018-096401-A-I00).

References

- Danckert, J. (2019). Boredom: Managing the Delicate Balance Between Exploration and Exploitation. En J. Ros Velasco (Ed.), *Boredom Is in Your Mind*, 37-53. Cham: Springer.
- Danckert, J. y Merrifield, C. (2018). Boredom, Sustained Attention and the Default Mode Network. *Experimental Brain Research* 236 (9), 2507-2518. doi: 10.1007/s00221-016-4617-5.
- Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, 7(5), 482-495.
- Farmer, R. y Sundberg, N. D. (1986). Boredom Proneness: The Development and Correlates of a New Scale. *Journal of Personality Assessment* 50 (1), 4-17. doi: 10.1207/s15327752jpa5001_2.
- FER Challenge.(n.d.).
<https://www.kaggle.com/Ashishpatel26/Facial-Expression-Recognitionferchallenge>.
- Finkielstein, M. (2020). Class-related academic boredom among university students: a qualitative research on boredom coping strategies. *Journal of Further and Higher Education*, 44(8), 1098-1113.
- Goetz, T., & Hall, N. C. (2014). Academic boredom. In *International handbook of emotions in education* (pp. 321-340). Routledge.
- Isacescu, J., Struk, A. A. y Danckert, J. (2017). Cognitive and Affective Predictors of Boredom Proneness. *Cognition and Emotion* 31 (8), 1741-1748. doi: 10.1080/02699931.2016.1259995.
- Keras API. (n.d.). <https://keras.io/Api/Preprocessing/Image/>.
- Kroes, S. (2007). Detection Boredom in Meetings. 3-7.
- Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: An investigation into the contributors, moderators and outcomes of boredom amongst university students. *British Educational Research Journal*, 35(2), 243-258.
- Merrifield, C., & Danckert, J. (2014). Characterizing the psychophysiological signature of boredom. *Experimental Brain Research*, 232(2), 481-491. <https://doi.org/10.1007/s00221-013-3755-2>
- Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary educational psychology*, 36(1), 49-59.
- O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. December. <http://arxiv.org/abs/1511.08458>
- Pekrun, R. (2005). Progress and open problems in educational emotion research. *Learning and Instruction*, 15(5), 497-506. <https://doi.org/10.1016/j.learninstruc.2005.07.014>
- Rebollo-Catalán, Á., García Pérez, R., Barragán Sánchez, R., Buzón García, O., & Vega Caro, L. (2008). Las emociones en el aprendizaje online. *RELIEVE - Revista Electronica de Investigacion y Evaluacion Educativa*, 14(1), 1-23. <https://doi.org/10.7203/relieve.14.1.4201>
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015, 815-823. <https://doi.org/10.1109/CVPR.2015.7298682>
- Sharp, J. G., Hemmings, B., Kay, R., Murphy, B., & Elliott, S. (2017). Academic boredom among students in higher education: A mixed-methods exploration of characteristics, contributors and consequences. *Journal of Further and Higher Education*, 41(5), 657-677.
- Sharp, J. G., Hemmings, B., Kay, R., & Sharp, J. C. (2019). Academic boredom and the perceived course experiences of final year Education Studies students at university. *Journal of Further and Higher Education*, 43(5), 601-627.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1-14
- Silvia, P. J. (2001). Interest and Interests: The Psychology of Constructive Capriciousness. *Review of General Psychology*, 5(3), 270-290. <https://doi.org/10.1037/1089-2680.5.3.270>
- Tze, V. M., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review*, 28(1), 119-144.
- Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The MAC model of boredom and cognitive engagement. *Psychological Review*, 125(5), 689-713. <https://doi.org/10.1037/rev0000097>
- Xie, J., Xu, J., Wei, T., Gallo, K., Giles, M. E., Zhan, Y., ... & Liu, X. (2021). Contributing Factors, Attribution, and Coping in Academic Boredom: An Exploratory Case Study of Graduate Students in Education. *Adult Learning*, 1045159520987304.
- Zhang, Q., Zhang, M., Chen, T., Sun, Z., Ma, Y., & Yu, B. (2019). Recent advances in convolutional neural network acceleration. *Neurocomputing*, 323, 37-51. <https://doi.org/10.1016/j.neucom.2018.09.038>
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1), 43-76. <https://doi.org/10.1109/JPROC.2020.3004555>

DIEGO ISAR MUÑOZ
ÁLVARO DAVID ORTIZ MARCHUT
DANIEL PADILLA RODRÍGUEZ
SOFÍA PRIETO IBÁÑEZ

Ult. actualización 15 de junio de 2021
T_EX lic. LPPL & powered by **TEFLON**

Esta obra está bajo una licencia Creative Commons “Reconocimiento 4.0 Internacional”.

